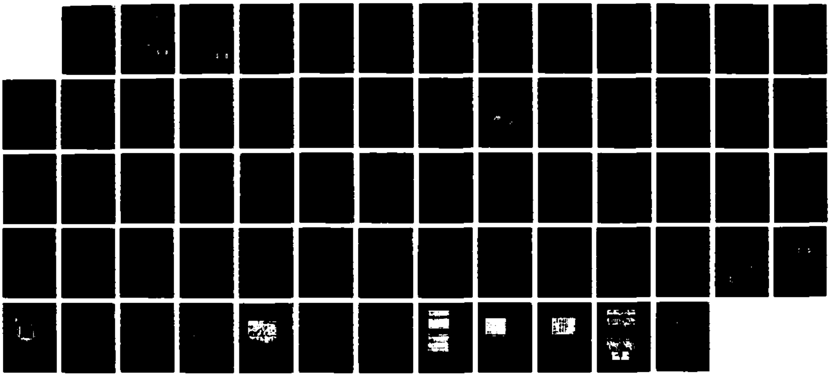
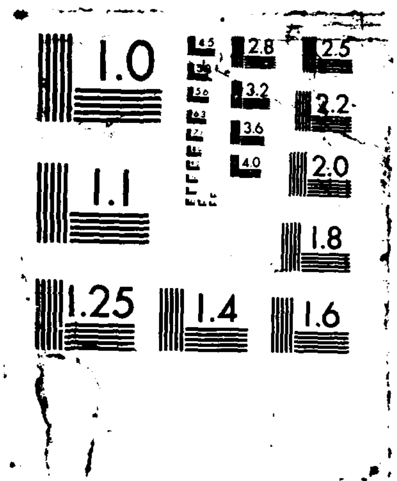
AD-A191 481 OPTICAL COMPUTING BASED ON NEURONAL MODELS(U) MOORE 1/1
SCHOOL OF ELECTRICAL ENGINEERINN PHILADELPHIA PA
ELECTRO-OPTICS AND MI CROWAVE-OPTICS LAB N H FARHAT
UNCLASSIFIED OCT 87 EO/MO-12 N00014-85-K-2036 F/G 23/3 NN




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

Final

(4)

DTIC FILE COPY

AD-A191 481

OPTICAL COMPUTING BASED ON
NEURONAL MODELS

Annual Report

DTIC
ELECTE
JAN 25 1988
S D
E

UNIVERSITY of PENNSYLVANIA

The Moore School of Electrical Engineering

PHILADELPHIA, PENNSYLVANIA 19104-6390

87 11 30 123

4

University of Pennsylvania
The Moore School of Electrical Engineering
200 South 33rd Street
Philadelphia, PA 19104-6390

OPTICAL COMPUTING BASED ON NEURONAL MODELS

Annual Report

Prepared by

Nabil H. Farhat

For
Naval Research Laboratory
4555 Overlook Avenue, S.W.
Washington, D.C. 20375
Attn: Code 5709

Under

Grant No. N00014-85-K-2036

October 1987

EO/MO Report No. 12
Period: 2/16/86 - 2/15/87

TABLE OF CONTENTS

REPORT DISTRIBUTION LIST

	Copies
Naval Research Laboratory 4555 Overlook Ave., S.W. Washington, D.C. 20375 Attn: Code 5709	1
ONRR National Academy of Sciences 2100 Pennsylvania Ave., N.W. Washington, D.C. 20037-3202	1
Director Naval Research Laboratory 3555 Overlook Ave., S.W. Washington, D.C. 20375 Attn: Code 2627	6
Defense Technical Information Center Bldg. 5, Cameron Station Alexandria, VA 22314	12
Program Management Office 1400 Wilson Blvd. Arlington, VA 22209-2308 Attn: Dr. John Neff	1

OPTICAL ASSOCIATIVE MEMORIES: FIRST STEP
TOWARDS NEUROMORPHIC OPTICAL DATA PROCESSING

1. INTRODUCTION

Ever since the fit between what neural net models can offer (collective, iterative, nonlinear, robust, and fault-tolerant approach to information processing) and the inherent capabilities of optics (parallelism and massive interconnectivity) was first pointed out and the first optical associative memory demonstrated in 1985 [1]~~,[2]~~ ~~(see~~ also Appendices I and ~~III of~~ this report~~)~~, work and interest in neuromorphic optical signal processing has been growing steadily. For example, work in optical associative memories is currently being conducted at several academic institutions (e.g., California Institute of Technology, University of Colorado, University of California - San Diego, Stanford University, University of Rochester, and the author's own institution the University of Pennsylvania) and at several industrial and governmental laboratories (e.g., Hughes Research Laboratories - Malibu, the Naval Research Laboratory, and the Jet Propulsion Laboratory). In these efforts, in addition to the vector marix multiplication with thresholding and feedback scheme utilized in early implementations, an arsenal of sophisticated optical tools such as holographic storage, phase conjugate optics, and wavefront modulation and mixing are being drawn upon to realize associative memory functions. Such functions include auto-associative, hetero-associative and sequential or cyclic storage and recall [3]-[8] with signal recovery from partial information receiving much attention as potential application [5],[9],[10].

It is gradually becoming clear however, that associative memory is only one apparent function of biological neural nets that lends itself to optical implementation. Optics can pay a useful role in the implementation of artificial neural nets capable of self-organization and learning i.e., self-programming nets. One can safely state that self-organization and learning is the most distinctive single feature that sets neuromorphic processing apart from other approaches to information processing. Learning in these nets is by adaptive modification of the weights of interconnections between neurons (plasticity). It can be supervised or unsupervised, deterministic

or stochastic. Work on optical learning networks is currently being pursued by two of the groups mentioned earlier, that at the University of Pennsylvania (Penn) and that at the California Institute of Technology (Caltech). Important progress in multilayered optical learning networks based on holographically inerconnected nonlinear Fabry-Perot etalons has been achieved by the Caltech group [11]. Making use of a deterministic error back-propagation algorithm these nets can learn the connectivity weights that represent associations they are presented with. The focus in this work is on the use of volume holograms formed in photo-refractive media, as opposed to planar holograms, for defining and storing the interconnectivity patterns between neurons. A clever fractal based method for the implementation of arbitrary interconnects input and output planes defined within the net's architecture with minimal cross-talk has also been devised and verified. The effort at Penn focuses on the other hand on architectures and methodologies for opto-electronic implementation of stochastic learning in self-organizing multilayered nets. This work is based on simulated annealing within the framework of a Boltzmann machine [12]. Stochastic rather than deterministic learning is of primary interest because of its physical plausibility and because it can shed light on the way nature has turned noise present in biological neural nets to work to its advantage. The primary result of this work to date is a scheme that combines optical random array generators with the parallelism of optics to accelerate stochastic learning by an estimated fator of 10^5 as compared to serial machine executions of the same learning algorithm. The scheme basically imparts to the neurons a random threshold component that produces controlled shaking of the "energy landscape" of the net which can, so to speak, shake the net loose whenever it tends to get stuck in a state of local energy minimum accelerating thereby its successful descent to the state of global energy minimum (or one very close to it) which is a requirement of the learning algorithm.

The above work on optical learning nets is helping bring into focus the acute need for suitable materials and devices for the implementation of programmable interconnects and plasticity. Examples are modifiable nonvolatile volume holographic materials, spatial light modulators, and dense arrays of nonlinear light amplifiers or optically bistable elements.

Biological neural nets were evolved in nature for one ultimate purpose: that of maintaining and enhancing survivability of the organism they reside in. Embedding artificial neural nets in man-made systems, and in particular autonomous systems, can serve to enhance their survivability and therefore reliability. Survivability is also a central issue in a variety of systems with complex behavior encountered in biology, economics, societal models, and military science. One can therefore expect neuromorphic processing to lay an increasing role in the modeling and study of such complex systems especially if optical techniques can be made to furnish speed and flexibility. Finally, one should expect also that software development for emulating neural functions on serial and parallel digital machines will not continue to be confined to the realm of straight-forward simulation, but spurred by the mounting inerest in neural processing, will move into the algorithmic domain where fast efficient algorithms are likely to be developed becoming to neural processing what the FFT was to the discrete Fourier transform. Thus we expect that advances in optical and digital neuromorphic signal processing will proceed in parallel.

2. RESEARCH ACCOMPLISHMENTS

During the period of this report, a comparative study of two equivalent implementations of the Hopfield model one employing direct storage and inner product recall and the other outer product storage and recall was conducted. The results of this study reveal clearly that the distinctive feature of neural net distributed processing as compared to straight forward template matching is self-organization and learning, which is not possible in inner product schemes because of the vanishing of the connectivity matrix T_{ij} in such schemes. Thus self-organization and learning is what sets neuromorphic processing apart from other conventional approaches to signal processing. It has the important advantage of alleviating the programming complexity of artificial neural nets (i.e. computing the T_{ij} matrix and setting the weights of interconnections accordingly). It should be noted that in serial computation programming complexity is not an issue however computational complexity of certain problems encountered in vision, remote sensing, and combinatorial optimization is. As the degree of parallelism of a computing machine increases, computational complexity decreases but is replaced by

increased difficulty of programming. Note that once the task of programming a neural net is completed a solution is found usually in a few time constants of the switching elements. That is the computational complexity virtually vanishes. By devising networks that can organize themselves (self-modify their synaptic weights) and hence learn or progrm themselves, the programming complexity of such nets is reduced drastically. Such nets can learn from examples presented to them and can capture the underlying structure of the entities presented. Learning can be deterministic (e.g. error backpropagation) or stochastic (simulated annealing in a Boltzmann machine framework), supervised or unsupervised. Accordingly in another aspect of our research, an architecture for the partitioning of an opto-electronic analog of a neural net into a multilayered structure suitable for use in the study and implementation of self-organization and stochastic learning has been developed (see Appendix II). Of particular interest in this work is stochastic learning by fast annealing in the context of a Boltzmann machine formalism.

The above work on optical learning nets is helping bring into focus the acute need for suitable materials and devices for the implementation of programmable interconnects and plasticity. Examples are modifiable nonvolatile volume holographic materials, spatial light modulators, and dense arrays of nonlinear light amplifiers or optically bistable elements. We have given attention in our work to magneto-optic spatial light modulators because of their nonvolatility and potential speed for use as modifiable synaptic mask or connectivity matrices. Although the switching time of a single pixel in such devices can be 100 nsec or less, the frame rate of commercially available units ranges between a few tens to a few hundred frames/sec depending on device size. A careful study of the addressing of such a device (see Appendix IV) reveals that parallel addressing schemes (one row at a time or one column at a time) can not be applied to exploit the fast pixel switching time in order to achieve high frame rates because of thermal and possible magnetic stress limitation. Ruling out special cooling arrangements which are bound to complicate operation and reduce the utility of the device we have developed a parallel/serial addressing scheme that avoids the above limitations and enables frame switching time of slightly less than a millisecond for a 48x48

pixel device (see Appendix IV for details). The scheme relies on the use of external magnetic field bias to reduce the amplitude and duration of the current pulse need to switch a pixel. Power dissipation is thus reduced to the allowable level of the device. A fast driver circuit capable of accepting 16 bit parallel input from a computer conroller (or a dedicated microprocessor) in 2μsec and driving 16 pixels of a row serially in 2μsec was devised and successfully operated. The serial addressing eliminates the possible magnetic stresses that arise in parallel addressing schemes. When used with a dedicated μprocessor the magneto-optic SLM is expected to be able to operate at 1000 frames/sec which would be particularly attractive in the implementation of neuromorphic optical learning machines in order to accelerate learning.

3. CONCLUSION

The deterministic iterative update equation for a neural net can be implemented in two equivalent opto-elecronic architectures, one involving outer product and the other inner product. Despite several attractive features, the inner product scheme appears to lack the ability of self-organization and learning and hence self-programmability where the net can find by itself what are the connectivity weights that best describe the entities it is given to learn. To this end we have shown that an opto-electronic analog of a neural net can be partitioned into layers of input, output, and hidden (internal, or buffer) neurons to facilitate the study and implementation of "optical learning machines". Furthermore the scheme enables partitioning a net into any desired number of layers with any prescribed communication pattern between them. Such nets appear to be well suited for implementing learning algorithms. In particular evidence is mounting in our work that opto-electronic techniques have much promise for implementing stochastic learning by simulated annealing in a context of Boltzmann machine where spatio-temporal optical noise can play an important role in accelerating the stochastic learning algorithms by several orders of magnitude as opposed to software implementation on serial computing machines. Such fast optical learning machines can be "softened" for the learning phase and then "hardened" to act as associative memory merely by controlling the noise level in the net. Furthermore synaptic modification

(plasticity) in these nets appear to be achievable by fast nonvolatile electronically addressed magneto-optic spatial light modulators. These are basically however binary transmittance SLM and hence require the development of binary learning rules. Search for such rules is underway in our work.

4. PUBLICATIONS AND OTHER ACTIVITIES

During the period of this report the following papers were published or presented at technical meetings:

- N.H. Farhat, "Architectures for Opto-Electronic Analogs of Self-Organizing Neural Networks", Topical Meeting on Optical Computing, Technical Digest Series 1987, Vol. 11 (Optical Society of America, Wash. D.C. 1987), pp. 125-128.
- N.H. Farhat, "Robust Signal Recovery and Recognition with Optical Analogs of Neural Nets and Spurious Memory Discrimination", SPIE, Vol. 700 IOCC, Proc. 1986 Intern. Opt. Comp. Conf., pp. 283-288.
- N.H. Farhat, "Neural Net Models and Optical Computing", in Optical Computing, H.H. Szu (Ed.), SPIE Vol. 634, pp. 307-309 (1987).
- N.H. Farhat, "Architectures and Implementation Concepts for Opto-Electronic Analogs of Self-Organizing Neural Networks", Presented at the Second Conference on Neural Networks for Computing, Snowbird, Utah April 1-4, 1987.

Also during the period of this report, N. Farhat attended the following conferences:

Conference on Current Topics in Neurobiology, Institute for Theoretical Physics, University of California, Santa Barbara, CA, Sept. 2-6, 1986.

Degrees Awarded:

S. Miyahara, Ph.D., University of Pennsylvania Dissertation: "Automated Radar Target Recognition Based on Models of Neural Nets", (1987).

Patent Disclosure:

Super-Resolution - Patent disclosure filed April 9, 1987 on behalf of the University of Pennsylvania, by University Patent Inc., 1465 Post Rd. East, P.O. Box 901, Westport, CT 06881

5. REFERENCES

1. D. Psaltis and N. Farhat, Opt. Lett. 10, (1985).

2. N.H. Farhat, D. Psaltis, A. Prate, and E. Paek, Appl. Opt. 24, 1469, (1985).

3. A.D. Fischer, C. Giles, and J. Lee, J. Opt. Soc. Am. A 1, 1337, (1984).

4. D.Z. Anderson, Opt. Lett. 11, 56, (1986).

5. B. Soffer, G. Dunnign, Y. Owechko, and E. Maron, Opt. Lett. 11, 118, (1986).

6. A. Yariv and S.K. Kwong, Opt. Lett. 11, 186, (1986).

7. B. Kosko and C. Guest, Proc. SPIE, 758, (in press).

8. M. Takeda and J.W. Goodman, App. Optics, 25, 3033, (1986).

9. N. Farhat, S. Miyahara, and K.S. Lee, in Neural Networks for Computing, J.S. Denker (Ed.), 146, Am. Inst. of Phys., New York, (1986).

10. Y.S. Abu-Mustafa and D. Psaltis, Scientific American, 256, 88 (1987).

11. K. Wagner and D. Psaltis, Proc. SPIE, 756 (in Press).

12. N. Farhat, Architectures for Opto-Electronic Analogs of Self-Organizing Neural Nets", Topical Meeting on Optical Computing, Technical Digest, Series 1987, Vol. 11, (Optical Society of America, Wash. D.C., 1987), pp. 125-128.

6. APPENDICES

I. Robust Signal Recovery and Recognition with Optical Analogs of Neural Networks and Spurious Memory Discrimination (Reprint).

II. Architectures for Opto-Electronic Analogs of Self-Organizing Neural Networks (Reprint).

III. Neural Net Models and Optical Computing - a Brief Overview (Reprint).

IV. Scheme for Enhancing the Frame Rate of Magneto-Optic Spatial Light Modulators.

Appendix I

Reprinted from SPIE Vol. 700—1986 International Optical Computing Conference

Robust signal recovery and recognition with optical analogs of neural nets and spurious memory discrimination

N.H. Farhat

Department of Electrical Engineering, University of Pennsylvania
200 South 33rd Street, Philadelphia, Pennsylvania 19104-6390

Abstract

Storage recipes used in forming associative content addressable memory (ACAM) based on models of neural nets are known to generate false spurious memory states together with those being intentionally stored. Effective use of such memories, and of recent optical embodiments of them, may require either exorcising the spurious states or devising a means for discriminating against them when they are evoked. In this paper an opto-electronic spurious memory discriminating circuit is described. It is used in conjunction with an ACAM, is designed to ignore spurious states when they occur, and indicates which of the nominal states is evoked by the stimulus achieving thereby robust recognition. The latter recognition circuit discussion is used as a vehicle for pointing out distinctive features of the "neural net" approach to signal processing as opposed to other approaches to performing the tasks described.

Introduction

The brain is regarded by many to be our last remaining frontier of exploration. With its more than 10^{10} neurons and estimated 10^{15} interconnections, where memory is believed to reside and to be stored electro-chemically as synaptic weights that are modified by learning and experience, the brain represents a formidably complex structure that does not lend itself readily to analysis and study. Never-the-less, steady advances in neuroscience, biology, and psychology during the last few decades are providing today such a level of understanding of the structural and functional properties of the brain as to allow simplified but non-the-less meaningful modeling of its neural networks. Several such models exist. These usually involve a large number of linear or nonlinear neuron-like element massively interconnected with each other through a modifiable connectivity or synaptic matrix where memories are stored in accordance to prescribed recipes with the "neural net" as a whole operating asynchronously with or without feedback loops[1-9]. The assumption of asynchronous operation stems from the lack of evidence of a "master clock" or controller in the brain, i.e., the assumption of an "egalitarian brain"[10]. Despite their highly simplified nature, these models exhibit, when implemented and exercised in either software or hardware form, striking similarities in their behavior to known capabilities of the brain. These capabilities include: content addressability and associative recall, robustness (tolerance to error and noise in the data it is presented with, and to element non-uniformity and failure), signal recovery from partial information, and a disconcerting, but fascinating, property of generalization, i.e., generation of false or spurious memory states in the course of loading (learning) desired ones.

The relationship between the number and composition of spurious states and the nominal memory states is not fully understood. It is known, however, that the number of spurious states increases sharply when the number M of entities stored or nominal memory states exceeds $N/4\ell nN$[39-41] where N is the number of neurons in the network. It has been hypothesized that REM (rapid eye movement) sleep and dreaming are means with which the brain rids itself of such false memories in order to stabilize real or nominal memory states[11]. Application of an "unlearning" algorithm[12] that is basically the reverse of the Hebbian learning process (see for example ref. 7), has been shown to reduce the number of spurious states and to increase accessibility of the nominal memory states. The unlearning process consists of stimulating the neural net model with random starting states and noting those final stable states that are not nominal memory states of the network, i.e., do not coincide with the stored states, for subsequent unlearning. Clearly a network of N binary (on-off) neurons requires 2^N trials to identify <u>all</u> of its false memory states before their unlearning can begin. A network of N=32 neurons for example would require $2^{32} \approx 4.3\ 10^9$ trials which is clearly impractical even for such a small network. Hence such a random trial approach to identifying false states and unlearning can not be effective in removing all false states but can merely reduce their number. Some false states will always be present. This raises a practical question: how could an associative memory containing false memory states intermingled with the nominal stored memory states be used effectively in practice for example in a robust recognition system? Before dealing with this question it is worth reviewing the role of optics in all of this.

The parallelism and massive connectivity features of neural nets have drawn attention recently to the feasibility of implementing associative memory and other functions exhibited by such networks efficiently by optical means as both of these features happen to be also the traditional main strengths of optics. The marriage of neural modeling and optics is attractive as it can add the power of iterative feedback and decision making (thresholding) to the traditional capabilities of optical processing. Several optical architectures and implementations involving a relatively small number of "neurons" but with potential for being extendable to much larger networks have recently been described[13-25].

The utility of such optical associative memory networks hinges however on the ability to identify and ignore the false memory states of the network when these are evoked by means other than the tedious and impractical unlearning procedure mentioned earlier. It is the purpose of this paper to describe a solution to this problem and to use the solution as a vehicle for pointing out the distinctive features of the neural net model approach to signal processing as opposed to other approaches that can equally perform similar functions.

The proposed solution is based on ignoring the spurious states when these are evoked by means of a recognition or identification "circuit" which is to be used in conjunction with the ACAM. The identification circuit, which is fed the output state of the ACAM, is designed to ignore all spurious or false states of the memory and to identify which of the nominal or real states of the memory has been evoked by the input stimulus. It has also the capability of reducing the dimensionality of the stimulus being identified from a single point in the N-dimensional state space of a network of N-neurons to a single descriptive statement or a word label. This process of reduction of dimensionality may be viewed as being analogous to the now out-dated notion of "grandmother cell"[26] in the brain where highly selective higher order neurons respond only and only when a particular distinctive feature is present in the stimulus. It is now believed,[6,27], that neural sensory and processing networks handle information in such a way as to achieve as complete a representation of the stimulus as possible with a spatio-temporal neural activity trace involving the smallest number of discharging neurons and not just a single firing cell.

For specific explanation of the discriminating "circuit" we refer to Figure 1 where the state vector $h_i^{(mo)}$ i=1,2...N of the ACAM represented by the photodetector array output as a unipolar binary vector N bits long serves as the input to the discriminating circuit. Note the ACAM itself (see for example ref. 13 or 14) is not shown in Figure 1. The vector $h_i^{(mo)}$ and its independently generated complement $\bar{h}_i^{(mo)}$ are compared independently in two optical inner-product vector matrix multipliers with the nominal state vectors $h_i^{(m)}$ i=1,2...N; m=1,2...M and their complements $\bar{h}_i^{(m)}$ stored as rows of the mask T and its complementary $\bar{T}$ respectively. By adding the outputs of the corresponding elements in the two photodetector arrays as shown, only one adder output will exceed N-ε where ε is a small number exceeding the noise level of the system.

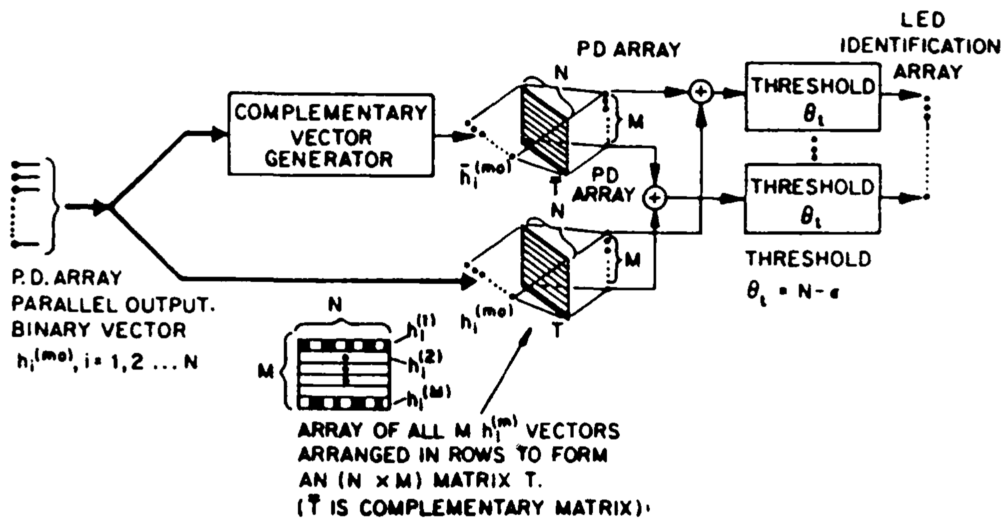


Figure 1. Spurious memory discriminator and recognition circuit.

Thresholding the adders outputs to N-ε will insure that only one output will be present in the LED identification array at the LED or terminal corresponding to the nominal memory state vector coinciding with $h_i^{(mo)}$. In this fashion an ACAM with a "grandmother-cell" recognition capability can be implemented.

Discussion and Conclusions

In the above scheme the ACAM performs a nearest-neighbor search based on an outer product storage recipe[7] while the recognition and discriminating circuit recognizes the ACAM output based on an inner-product correlation scheme followed by a maximum value discriminator (thresholding and labeling). The nearest-neighbor search operation can however be performed by the recognition and discriminating circuit directly without resort to the ACAM. What does the neural net approach then offer that conventional signal processing and computing methods do not? The answer to this question may be found in the following observations which were also given elsewhere,[37,38].

(a) Neural nets and their models provide us with a new way of viewing signal processing and computing problems that may lead to solutions and applications not thought of otherwise in a manner very much reminiscent to what happened with the advent of holography several decades ago.

(b) The brain and its neural organization are the results of a prolonged evolutionary process in which only those permutations that enhanced the survivability, and by implication the function, of the organism have been retained and all other permutations have been discarded through "survival of the fittest" process. The result is a biological "computer" that has no equal at present among artificial systems when it comes to tasks of recognition, classification, categorization, generalization, and perhaps optimization. It is not surprising therefore to see the current interest in brain mechanisms in the artificial intelligence, cognitive science, and pattern recognition communities. Even early researchers in computer science were concerned with brain-related attributes of computing systems[28]. A serial approach to computing was however adopted and pursued because of better understanding, easier mathematical modeling, and because a collective or parallel approach to computation would have been technologically and economically not feasible at the time. There is much therefore that we can afford to learn from the brain and its neural nets that can prove to be useful in artificial man made systems today. The general idea here would be not to attempt to build systems that try to emulate the brain, as this would be unrealistic, but to gain insights in its operation and mechanisms, as to allow us to be able to glean information about those attributes and functions that would be worth incorporating in man made systems. Optical analogs of neural nets can be a useful tool in gaining such insights and as a way of ultimately simulating neural nets consisting of thousands to millions of elements.

(c) Information processing in the brain can be characterized as collective, adaptive, highly nonlinear, and relying heavily on feedback. We know that all these are attributes of powerful signal processing algorithms. What is amazing is that these capabilities are achieved in networks that appear to be homogeneous in their general structure i.e., only a few different types of neurons (cells) are involved and the process followed by most neurons is macroscopically similar in the sense that each neuron receives inputs from other neighboring neuron and decides to change its state or not depending on the nature of the inputs and the synaptic weights they make with the neuron where memory is stored. In other words the brain exhibits a fractal (self-similar) feature in its structure. It is now recognized that the same basic structure of highly interconnected neurons can be involved in the formation of signal-adaptive self-organized networks for unsupervised learning and feature extraction from sensory data[4,29], in the formation of associative memory modules tuned to respond to specific features of the sensory data,[30-34], and for recognition and perception "circuits" of the brain. This is a tremendous flexibility that is achieved with little apparent specific design unlike conventional circuit design where function specific components and parts are connected in accordance to specific rules to perform required signal processing tasks. As evidence of this flexibility it is worth noting in this regard that in associative memory based on neural net models the function of memory and signal processing involved in recall are totally intermingled.

(d) The fractal nature of neural nets, their robustness and fault tolerance can be immensely useful for modern VLSI technology. Continuing advances in microfabrication, and optical technologies promise to make it possible to fabricate large numbers of massively interconnected decision making switching elements with low power consumption. Neural net models and architectures can provide such structures (possibly in the form of opto-electronic chips) with the robustness and the fault-tolerance badly needed in VLSI to alleviate the central problem of yield. Less than perfect chips may no longer be discarded as rejects but can find use in artificial neural net processors or computers.

(e) Architectures and Optical implementations of 2-D neural net arrangements can provide content addressable associative memory modules that are suited for use with 2-D pictorial data. In practice such pictorial data would be, in the form of distortionless descriptors, (scale, rotation, and size invariant classifiers or feature spaces) derived from sensory array data generated for example by electromagnetic or acoustic (including seismic and sonar) receiver arrays. Associative memory modules can tell us then rapidly, in a matter of a few time constants of the neuron-like elements used, which member of an ensemble of input descriptors is recognized by the memory even when the descriptor information it is presented with happens to be sketchy (incomplete or noisy). Recognition of an input is manifested by the neural net converging and settling into one of its nominal states corresponding to that stored entity which most resembles the input. It is clear that this operation by itself may not be useful unless some means for identifying or recognizing the outcome of the nearest neighbor search is incorporated as described earlier. This also may suggest that in the absence of such "recognizer", auto-associative memories may be intended as building blocks of more complex processors in which higher order hierarchical processing takes place, in a manner not yet fully understood. Such processing would have the aim of first reducing the information content of the signals flowing concurrently through such a network of associative CAMs and then fusing their outcomes into a single meaningful concept or perception leading to subsequent action as shown schematically in Figure 2 which depicts a generalized smart sensing and recognition system.

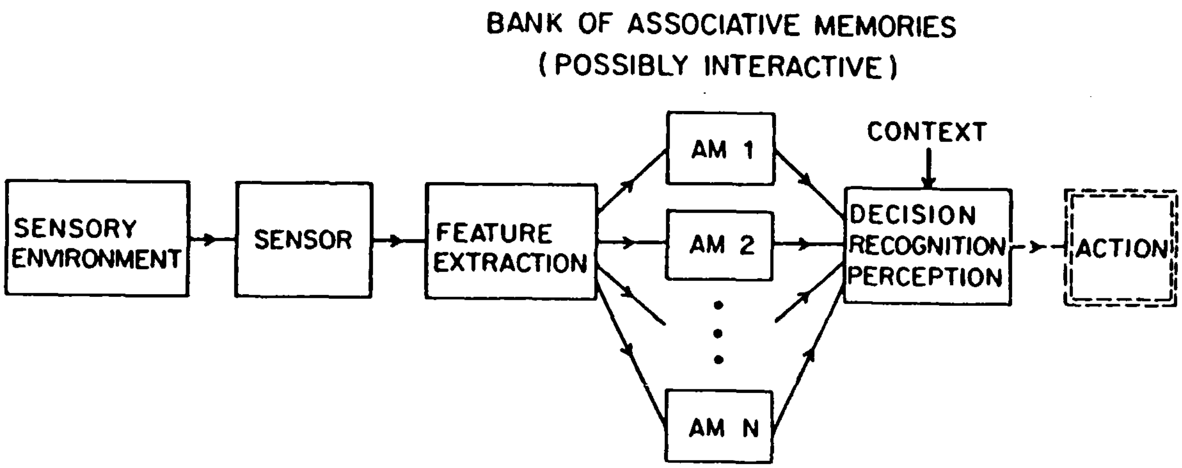


Figure 2. Concept of a smart sensing and automated recognition system employing self-organized feature extraction and array of associative or content addressable memories based on models of the neural nets.

In biological systems, such perception is thought to be associated with a prescribed trace of neural activity (spatio-temporal firing pattern of neurons) in the cortex. Associative memory modules by themselves become more meaningful when used in a hetero-associative mode where the learning set of descriptors or classifiers is not stored by auto-association but by heter-association with reference entities that are more easy to recognize i.e., reference entities that are tailored to the "recognizer" such as word labels or pictorial representations in the case of a human observer or coded analog or digital outputs for activating motor functions in robotic and artificial intelligence systems. Hetero-associative storage and recall of partial information can play an important role in smart sensing and automated recognition systems as was recently demonstrated with an example of identifying different types of aircraft from partial information about their sinogram classifiers[35] and individual faces from edge enhanced versions as classifiers[36]. In these neural net analogs the function of memory, processing, and recognition or labeling are carried out simultaneously by the same network of inter-connected "neurons". This is a distinctive feature of signal processing in such networks that sets them apart from other more familiar approaches to signal processing. The practical utility and implications of this unique feature remain to be seen. The intellectual challenge of understanding the dynamics and properties of neural nets and other variants of collective processing insure however that answers to these questions are forthcoming.

Acknowledgement

This work was supported by the Defense Advanced Research Projects Agency via the Naval Research Laboratory.

References

1. Widrow, B. and M.E. Hoff, "Adaptive Switching Circuits," WESCON Convention Record, Part IV, pp. 96-104. 1960. Also in Self-Organizing Systems, M.C. Yovitz, et. al. (Eds.) Spartan 1962.
2. Nakano, K., "Associatron-A Model of Associative Memory," IEEE Trans. on Systems Man & Cybernetics., Vol. SMC-2, pp. 380-388. 1972.
3. Kohonen, T., "Correlation Matrix Memories," IEEE Trans. on Computers, Vol. C-21, pp. 353-359. 1972.
4. Kohonen, T., Associative Memory, Springer Verlag, Heidelberg (1978). Also "Self-Organization and Associative Memory," Springer Verlag. 1984.
5. Willshow, D.J., "A Simple Network Capable of Inductive Generalization," Proc. Roy. Soc. London, Vol. 182, pp. 233-247. 1972.
6. Anderson, J.A., et. al., "Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model," Physiological Review, Vol. 34, pp. 413-451. 1977.
7. Hopfield, J.J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," Proc. Natl. Acad. Sci., USA, Vol. 79, pp. 2554-2558. 1982. Also, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons," Proc. Natl. Acad. Sci., USA, Vol. 81. 1984.
8. Grossberg, S., Studies of Mind and Brain, R. Reidel Pub. Co., Boston. 1982.
9. Sanderson, A.C., and Y.Y. Zeevi, (Eds.), Special Issue on Neural and Sensory Information Processing, IEEE Trans. on Syst. Man and Cybernetics, Vol. SMC-13. 1983.
10. Ferry, G., "The Egaliterian Brain," New Scientist, Vol. 109. 1986. Article based on the J. Hughlings-Jackson commemorative symposium "Hierarchies in the Brain," held at the London Hospital. 1985.
11. Crick, F.C. and G. Mitchison, "The Function of Dream Sleep," Nature, Vol. 304, pp. 111-114. 1983.
12. Hopfield, J.J., et. al., "Unlearning Has Stabilizing Effect in Collective Memories," Nature, Vol. 304, pp. 158-159. 1983.
13. Psaltis, D. and N. Farhat, "A New Approach to Optical Information Processing Based on the Hopfield Model," Digest of the 13th Congress of Intern. Commiss. on Optics, ICO-13, Saporro, Japan. 1984. Also, "Optical Information Processing Based on Associative-Memory Model of Neural Nets with Thresholding and Feedback," Opt. Lett., Vol. 10, pp. 98-100. 1985.
14. Farhat, N.H., et. al., "Optical Implementation of the Hopfield Model," App. Optics, Vol. 24, pp. 1469-1475. 1985.
15. Farhat, N.H. and D. Psaltis, "Architectures for Optical Implementation of 2-D Content Addressable Memories," Digest OSA Annual Meeting, Wash., D.C., p. 58. 1985.
16. Lee, K.S. and N.H. Farhat, "Content Addressable Memory with Smooth Transition and Adaptive Thresholding," Digest OSA Annual Meeting, Wash., D.C., p. 48. 1985.
17. Psaltis, D., et. al., "Acousto-Optic Implementation of Neural Network Models," Digest OSA Annual Meeting, Wash. D.C., p. 58. 1985.
18. Condon, D.J., et. al., "Optical Window Addressable Memory," OSA Topical Meeting on Optical Computing, Incline Villege, (Post deadline paper). 1985.
19. Fisher, A.D., C.L. Giles, and J.N. Lee, "Associative Processor Architectures for Optical Computing," J. Opt. Soc. Am., A, Vol. 1, pp. 1337. 1984. Also, "An Adaptive, Associative Optical Computing Element," Technical Digest, OSA Topical Meeting on Optical Computing, Incline Village, Nev., pp. WB4-1 to WB4-4. 1985.
20. Fisher, A.D., and C.L. Giles, "Optical Adaptive Associative Computer Architectures," Proc. of the IEEE COPCOM Spring Meeting, IEEE Computer Society, IEEE Cat. No. CH2135-2/85/0000/0342, pp. 342-344. 1985.
21. Fisher, A.D., et. al., "Implementation of Adaptive Associative Optical Computing Elements," SPIE, Vol. 625, Bellingham, WA. 1986.
22. Sofer, B., et. al., "Associative Holographic Memory with Feedback, Using Phase-Conjugate Mirrors," Opt. Lett., Vol. 11, pp. 118-120. 1986.
23. Anderson, D.Z., "Coherent Optical Eigenstate Memory," Opt. Lett. Vol. 11, pp. 56-58. 1986.
24. Yariv, A. and S.K. Kwong, "Associative Memories Based on Message-Bearing Optical Modes in Phase-Conjugate Resonator," Opt. Lett., Vol. 11, pp. 186-183. 1986.
25. Psaltis, D., et. al., "Shift Invariance in Optical Associative Memories," SPIE, Vol. 625, Bellingham, WA. 1986.
26. Restak, R.M., The Brain, Bantam Books, p. 54. 1984.
27. Barlow, H.B., "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology," Perception, Vol. 1, pp. 371-394. 1971.
28. Von Neuman, J., Probabilistic Logics and Synthesis of Reliable Organism From Unreliable Components in Automata Studies, C.E. Shannon and J. McCarthy (Eds.), Princeton University Press, Princeton, N.J., pp. 48-98. 1956.
29. Fukushima, K., "Neocognition: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," Biol., Cybern., Vol. 36, pp. 193-202. 1980.
30. Hubble, D.H., "The Visual Cortex of the Brain," Scientific American, Vol. 209, pp. 54-63. 1963.

31. Hubble, D.H. and T.W. Weisel, "Functional Architecture of Macaque Monkey Visual Cortex," Proc. Roy. Soc. Lond. B., Vol. 198, pp. 1-59. 1977.
32. Tootel, R.B., et. al., "Spatial Frequency Columns in Primary Visual Cortex," Science, Vol. 214, pp. 813-815. 1981.
33. Orban, G.A., Neuronal Operations in the Visual Cortex, Springer-Verlag, New York, see for example, sections 7.3.4 and 15.1.3. 1984.
34. Jones, J.P., and L.A. Palmer, "Simple Receptive Fields of Cat Striate Cortex: A Comparison with Gabor Functions in Two Dimensions of Space and Two Dimensions of Spatial Frequency," University of Pennsylvania - private communication. 1985.
35. Farhat, N., et. al., "Optical Implementation of 2-D Neural Nets and Their Application in Recognition of Radar Target," Presented at the Neural Networks for Computing Conference, Snowbird, Utah. 1986.
36. Psaltis, D., et. al., "Nonlinear Discriminant Functions and Associative Memories," Presented at the Neural Networks for Computing Conference, Snowbird, Utah. 1986.
37. Farhat, N.H., in "Optical and Hybrid Computing," Proceedings of Special Institute of the SPIE, H. Szu (Ed.) (to be published).
38. Farhat, N.H. and D. Psaltis, "Optical Implementations of Associative Memory Based on Models of Neural Nets," in Optical Computing, J. Horner (Ed.), (to be published).
39. Abu-Mustafa, Y.S. and J.M. St. Jacques, "Information Capacity of the Hopfield Model," IEEE Trans. on Inf. Theo., Vol. IT-31, pp. 461-464. 1985.
40. McElice, R.J., et. al., "Capacity of the Hopfield Associative Memory," Caltech Report. 1986.
41. Venkatesh, S. and D. Psaltis, "Information Storage and Retrieval in Two Associative Nets," Submitted to IEEE Trans. on Info. Theo.

Appendix II

ARCHITECTURES FOR OPTO-ELECTRONIC ANALOGS OF SELF-ORGANIZING NEURAL NETWORKS

Nabil H. Farhat
University of Pennsylvania
Electrical Engineering Department
Electro-Optics and Microwave-Optics Laboratory
Philadelphia, PA 19104-6390

Abstract

Architectures for partitioning opto-electronic analogs of neural nets into input/output and internal units to enable self-organization and learning where a net can form its own internal representations of the "environment" are described.

1. INTRODUCTION: In our preceeding work on optical analogs of neural nets, [1],[2], the nets described were programmed to do a specific computational task, namely a nearest neighbor search by finding the stored entity that is closest to the address in the Hamming sense. As such the net acted as a content addressable associative memory. The programming was done by computing first the interconnectivity matrix using an outer-product recipe given the entities we wished the net to store and become familiar with followed by setting the weights of synaptic interconnections or links between neurons accordingly.

In this paper we are concerned with architectures for opto-electronic implementation of neural nets that are able to program or organize themselves under supervised conditions, i.e., of nets that are capable of (a) computing the interconnectivity matrix for the associations they are to learn, and (b) of changing the weights of the links between their neurons accordingly. Such self-organizing networks have therefore the ability to form and store their own internal representations of the entities or associations they are presented with.

Multi-layered self-programming nets have been described recently [3]-[5] where the net is partitioned into three groups. Two are groups of visible or external input/output units or neurons that interface with the outside world i.e., with the net environment. The third is a group of hidden or internal units that separates the input and output units and participates in the process of forming internal representations of the associations the net is presented with, as for example by "clamping" or fixing the states of the input and output neurons to the desired associations and letting the net run through its learning algorithm to arrive ultimately at a specific set of synaptic weights or links between the neurons that capture the underlying structure of all the associations presented to the net. The hidden units or neurons prevent the input and output units from communicating with each other directly. In other words no neuron or unit in the input group is linked directly to a neuron in the output group and vice-versa. Any such communication must be carried out via the hidden units. Neurons within the input group can communicate with each other and with hidden units and the same is true for neurons in the output group. Neurons in the hidden group can not communicate with each other. They can only communicate with neurons in the input and output groups as stated earlier.

Two adaptive learning procedures in such partitioned nets have attracted considerable attention. One is stochastic involving a **simulated annealing** process [6],[7] and the other is deterministic involving an **error back-propogation** process [4]. There is general agreement however; that because of their iterative nature, serial digital computation of the links with these algorithms is very time consuming. A faster means for carrying out the required computations is needed. Never-the-less the work mentioned represents a milestone in that it opens the way for powerful collective computations in multilayered neural nets and in that it dispels earlier reservations [8] about the capabilities of early models of neural nets such as the Perceptron [9] when the partitioning concept is introduced. What is most significant and noteworthy, in our opinion, is the ability to now define buffered input and output groups with unequal number of neurons in a net which was not possible with earlier nets where all neurons participate in defining the initial (input) and final (output) states of the net.

2. ANALOG IMPLEMENTATIONS: Optics and opto-electronic architectures and techniques can play an important role in the study and implementation of self-programming networks and in speeding-up the execution of learning algorithms. We have done some exploratory work in this regard to see how the neurons in an opto-electronic analog of a neural net can be partitioned into groups with specific interconnection patterns. Here, for example, a method for partitioning an opto-electronic analog of a neural net into input, output, and internal units with the selective communication pattern described earlier to enable, stochastic learning, i.e., carrying out a simulated annealing learning algorithm in the context of a **Boltzmann machine** formalism is described. (see Fig. 1(a)). The arrangement shown in Fig. 1(a) derives from the neural network analogs we described earlier [2]. The

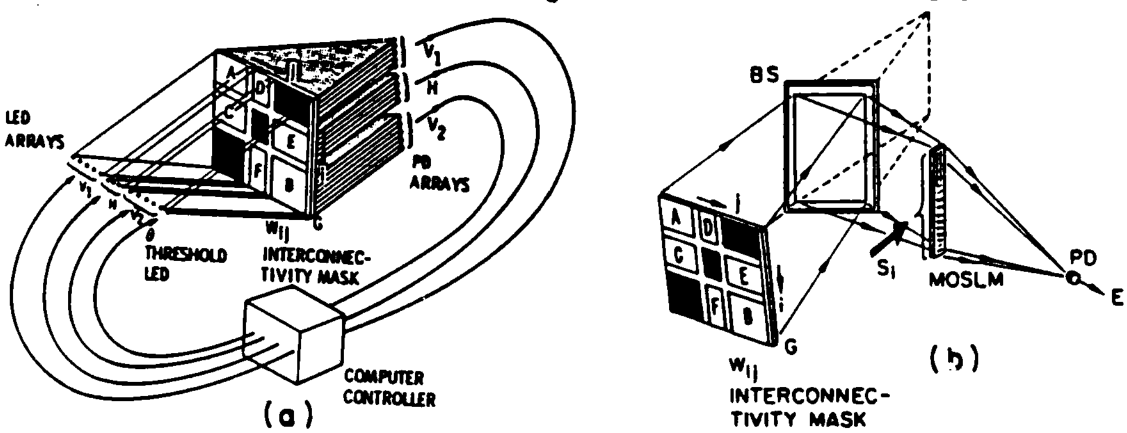


Fig. 1. Partitioning concept (a) and method for rapid determination of the net's energy E.

network, consisting of say N neurons, is partitioned into three groups. Two groups, V_1 and V_2, represent visible or exterior units that can be used as input and output units respectively. The third group H are hidden or internal units. The partition is such that $N_1+N_2+N_3=N$ where subscripts 1,2,3 on N refer to the number of neurons in the V_1, V_2 and H groups respectively. The interconnectivity matrix, designated here as W_{ij}, is partitioned into nine submatrices, A,B,C,D,E, and F plus three zero matrices shown as blackened or opague regions of the W_{ij} mask. The LED array

represents the state of the neurons, assumed to be unipolar binary LED on = neuron firing, LED off = neuron not-firing. The W_{ij} mask represents the strengths of interconnection between neurons in a manner similar to earlier arrangements [2]. Light from the LEDs is smeared vertically over the W_{ij} mask with the aid of an anamorphic lens system (not shown in Fig. 1(a)) and light emerging from rows of the mask is focused with the aid of another anamorphic lens system (also not shown) onto elements of the photodetector (PD) array. Also we assume the same scheme utilized in [2] for realizing bipolar values of W_{ij} in incoherent light is adopted here, namely by separating each row of the W_{ij} mask into two subrows and assigning positive values W^{+}_{ij} to one subrow and negative values W^{-}_{ij} to the other, then focusing light emerging from the two subrows separately onto pairs of adjacent photosite connected in opposition in the V_1, V_2 and H segment of the photodetector array. Submatrix A with $N_1 \times N_1$ elements, provides the interconnection weights of units or neurons within group V_1. Submatrix B with $N_2 \times N_2$ elements, provides the interconnection weights of units within V_2. Submatrices C (of $N_1 \times N_3$ elements) and D (of $N_3 \times N_1$ elements) provide the interconnection weights between units of V_1 and H and submatrices E (of $N_2 \times N_3$ elements) and F (of $N_3 \times N_2$) provide the interconnection weights of units of V_2 and H. Units in V_1 and V_2 can not communicate with each other directly because locations of their interconnectivity weights in the W_{ij} matrix or mask are blocked out (blackened lower left and top right portion of W_{ij}). Similarly units within H do not communicate with each other because locations of their interconnectivity weights in the W_{ij} mask are also blocked out (center blackened square of W_{ij}). The LED element θ is always on to provide a fixed or adaptive threshold level to all other units by contributing to the light focused onto only negative photosites of the photodetector (PD) arrays.

By using a computer controlled nonvolatile spatial light modulator to implement the W_{ij} mask in Fig. 1(a) and including a computer/controller as shown the scheme can be made self-programming with ability to modify the weights of synaptic links between its neurons to form internal representations of the associations or patterns presented to it. This is done by fixing or clamping the states of the V_1 (input) and V_2 (output) groups to each of the associations we want the net to learn and by repeated application of the simulated annealing procedure with Boltzmann, or other, stochastic state update rule and collection of statistics on the states of the neurons at the end of each run when the net reaches thermodynamic equilibrium.

For each clamping of the V_1 and V_2 units to one of the associations, annealing is applied, starting from an arbitrary W_{ij}, with switching states of units in H until thermodynamic equilibrium is reached. The state vector of the entire net, which represents a state of global energy minimum, is then stored by the computer. This procedure is repeated for each

association several times recording the final state vectors every time. The probabilities P_{ij} of finding the i-th and j-th neurons in the same state are then obtained. Next with the output units V_2 unclamped to let them free run like the H units the above procedure is repeated for the same number of annealings as before and the probabilities P'_{ij} are obtained. The weights W_{ij} are then incremented by $\Delta W_{ij} = \eta(P_{ij}-P'_{ij})$ where η is a constant that controls the speed and efficacy of learning. Starting from the new W_{ij} the above procedure is repeated until a steady W_{ij} is reached at which time the learning procedure is complete. Learning by simulated annealing requires calculating the energy E of the net [3],[5]. A simplified version of a rapid scheme for obtaining E opto-electronically is shown in Fig. 1(b). A slight variation of this scheme that can deal with the bipolar nature of W_{ij} would actually be utilized. This is not detailed here because of space limitation.

3. REMARKS: The partitioning architecture described is extendable to multilayered nets of more than three layers and to 2-D arrangement of neurons. Learning algorithms in such layered nets lead to multivalued W_{ij}. Therefore high-speed computer controlled SLMs with graded pixel response are called for. Methods of reducing the dynamic range of W_{ij} or for allowing the use of W_{ij} with ternary weights are however under study to enable the use of commercially available nonvolatile SLM devices that are mostly binary e.g., Litton's MOSLM.

4. ACKNOWLEDGEMENT: The work reported was supported by grants from DARPA/NRL, The Army Research Office, and the University of Pennsylvania Laboratory for Research on the Structure of Matter.

REFERENCES:

1. Psaltis, D. and N. Farhat, Digest of the 13-th Congress of Intern. Commiss. on Optics, ICO-13, Saporro, Japan, 1984. Also, Opt. Lett., Vol. 10, pp. 98-100, 1985.
2. Farhat, N.H., et. al., App. Optics, Vol. 24, pp. 1469-1475, 1985.
3. G.F. Hinton, et. al., Carnegie-Mellon University Report No. CMU-CS-84-119, May (1984).
4. D.E. Rumelhart, G.E. Hinton and R.J. Williams, Institute for Cognitive Science Report 8506, Univ. of California, San Diego, Sept. 1985.
5. T.J. Sejnowski and C.R. Rosenberg, Johns Hopkins University, Electrical Engineering and Computer Science Technical Report No. JHU/EECS-96/01 (1986).
6. N. Metropolis, et. al., J. Chem. Phys., Vol. 21, 6, pp. 1087-1092, June, 1953.
7. S. Kirkpatric, et. al., Science, Vol. 220, pp. 671-680, May 1983.
8. M.L. Minsky and S. Papert, Perceptrons, MIT Press, Cambridge, MA, (1969).
9. F. Rosenblatt, Principles of Neuro-Dynamics: Perceptions and the Theory of Brain Mechanisms, Spartan Books, Wash. D.C., 1962.

Appendix III

NEURAL NET MODELS AND OPTICAL COMPUTING: A BRIEF OVERVIEW

Nabil H. Farhat

University of Pennsylvania
The Moore School of Electrical Engineering
Electro-Optics and Microwave-Optics Laboratory
200 S. 33rd Street
Philadelphia, PA 19104-6390

ABSTRACT

A brief overview of background and developments in the emerging field of neural net models for optical computing is presented. At this early stage the field is offering a new and intellectually stimulating approach to signal processing that dove-tails with and compliments the capabilities of optics.

1. INTRODUCTION

Interest in neural network models (see for example, [1]-[9]) and their opto-electronic analogs stems from well recognized capabilities of the brain and the fit between what optics can do and what even simplified models of neural nets can offer toward the development of new approaches to collective signal processing. We know that the brain is not as good in arithmetic operations as a digital computer but when it comes to operations such as association, categorization, generalization, classifiction, feature extraction, recognition, and optimization it outperforms even the most powerful of todays computers. The brain's amazing capabilities in analyzing sensory data and in controlling motor function, let alone complex thought and intelligent reasoning, makes it an intriguing model for smart sensing and automated recognition systems and for robotic and automatic control systems with unescapable ramifiction for pattern recognition, artificial intelligence and autonomeous systems. An interesting aspect of its ability in processing sensory data is the ease with which it solves computationally complex problems associated for example with vision that are basically inverse problems [10]. These are computationally vexing because of their ill-posedness [11]. The brain's associative memory capabilities where nearest neighbor searches are performed successfully even when the information it is presented with is sketchy are evidence of its remarkable robustness where high levels of missing or erroneous data in the input can be tolerated. The ability of the brain to supplement missing information has exciting implications for super-resolution and other similar problems of signal recovery from incomplete and noisy data [11],[12]. The capabilities of the brain in rapid solution of optimization problems [13] are also well appreciated. Add to the above that the brain is amazingly fault tolerant as

its (cells or neurons), unlike cells in other parts of the body, are not renewable. Yet, despite loss of a non-negligible number of cells by the time one of us reaches the age of 50 we continue to function normally. We know something about how the brain processes information. We know that it does that in parallel or more precisely collectively by means of massively interconnected networks of neurons. There are anywhere between 10^{10}- 10^{11} neurons in the brain, each making about 10^{3}- 10^{4} synaptic interconnections with neighboring neurons for a total of 10^{13} to 10^{15} interconnections. Even when we assume the neuron to take only two states: firing or not firing i.e., binary neuron, the total number of degrees of freedom of the the brain is truly astronomical reaching $2^{10^{15}}$. Each neuron independently evaluates its state and decides to change it or not depending on whether the sum of its synaptic (exitatory and inhibitory) inputs exceeds a given threshold or not performing thus a highly nonlinear (logic) operation.

Massive connectivity and parallelism are the two main attributes of optics. Optics can therefore play an important role in the implementation of models of neural nets for computing and signal processing. Besides developments in programmable nonvolatile spatial light modulators [14], optical light amplifiers [15], and optical bistability devices [16] promise to play a useful role in the implementations of programmable connectivity matrices, and optical decision making elements leading ultimately to powerful neural net type processors.

2. CURRENT RESEARCH

The above observations have served as motivation for several workers [17]-[31] to investigate the feasibility and capabilities of optical implementations of neural nets. Optical implementations of 1-D and 2-D distributions of neurons (neural nets) have been considered and/or studied and evaluated in coherent and incoherent light. The performance of such networks is found to conform to theoretical prediction of storage capacity (number of entities that can be reliably stored) [32]-[34], and exhibits robustness and fault tolerance. In an optical implementation of an associative memory of N=32 neurons, with three stored entities errors of up to 30% in the input stimulus was tolerated during complete and correct recall of the nearest neighbor stored entity. Furthermore, accidental failure of about 10% of the neurons hardly caused noticable degradation in performance. Some of the above studies have been aimed at finding means of simplifying optical implementation of large neural nets, increasing the storage capacity, and finding ways for their interconnections as modules into more complex systems to perform higher order hierarchical processing.

It is worth noting that operations such as associative recall and nearest neighbor search performed in the above systems can be carried out by more conventional means without resort to neural net processing. What does the neural net approach then offer that conventional signal processing and computing methods do not?. The answer to this question may be found in the following observations: (a) Neural nets and their models provide us with a new way of viewing signal processing and computing problems that may lead to solutions and applications not thought of otherwise in a manner very much reminiscent to what happened with the advent of holography several decades ago. (b) The brain and its neural organization are the results of a

prolonged evolutionary process in which only those permutations that enhanced the survivability, and by implication the function, of the organism have been retained and all other permutations have been discarded through "survival of the fittest" process. The result is a biological "computer" that has, as pointed out earlier, no equal at present among artificial systems when it comes to tasks of recognition, classification, categorization, generalization, recognition, and optimization. It is not surprising therefore to see the current interest in brain mechanisms in the artificial intelligence, cognitive science, and pattern recognition communities. Even early researchers in computer science were concerned with brain-related attributes of computing systems [35]. A serial approach to computing was however adopted and pursued because of better understanding, easier mathematical modeling, and possibly because a collective or parallel approach to computation would have been technologically and economically not feasible at the time. There is much therefore that we can afford to learn from the brain and its neural nets that can prove to be useful in artificial man made systems. The general idea here is not to attempt to build systems that fully emulate the brain in all its functions, as this would be unrealistic, but to gain insights in its operation and mechanisms, as to allow us to be able to glean information about these attributes and functions that would be worth incorporating in man made systems. Optical analogs of neural nets can be a useful tool in gaining such insights and as a way of ultimately simulting neural nets consisting of thousands to millions of elements. (c) Information processing in the brain can be characterized as collective, adaptive, highly nonlinear and rely heavily on feedback. We know that all these are attributes of powerful signal processing algorithms. What is amazing is that these capabilities are

achieved in networks that appear to be similar homogeneous in their general structure i.e., only a few different types of neurons (cells) are involved and the process followed by most neurons is macroscopically similar in the sense that each neuron receives inputs from other neighboring neuron and decides to change its state or not depending on the nature of the inputs and the synaptic weights they make with the neuron where memory is stored. In other words the brain exhibits a fractal (self-similar) feature in its organization. It is now recognized that the same basic structure of highly interconnected neurons can be involved in the formation of signal-adaptive self-organized networks for unsupervised learning and feature extraction from sensory data [4],[36], in the formation of associative memory modules tuned to respond to specific features of the sensory data, [37]-[41] and for recognition and perception "circuits" of the brain. This is a tremendous flexibility that is achieved with little apparent specific design unlike conventional circuit design where function specific components and parts are connected in accordance to specific rules to perform required signal processing tasks. As evidence of this flexibility it is worth noting in this regard that in associative memory based on neural net models the function of memory and signal processing involved in recall are totally intermingled. (d) The fractal nature of neural nets, their robustnewss and fault tolerance can be immensely useful for modern VLSI technology. Continuing advances in microfabrication, and optical technologies promise to make it possible to fabricate large numbers of massively interconnected decision making switching elements with low power consumption. Neural net models and architectures can provide such structures (possibly in the form of opto-electronic chips) with the robustness and the fault-tolerance badly needed in VLSI to alleviate the central problem of yield. Imperfect chips may no

longer be regarded as rejects but can find use in artificial neural net processors or computers. (e) Architectures and Optical implementations of 2-D neural net arangements can provide content addressable associative memory modules that are suited for use with 2-D pictorial data. In practice such pictorial data would be, in the form of distortionless descriptors, (scale, rotation, and size invariant classifiers or feature spaces) derived from sensory array data generated for example by electromagnetic or acoustic (including seismic and sonar) receiver arrays. Associative memory modules can tell us then rapidly, in a matter of a few time constants of the neuron-like elements used, which member of an ensemble of input descriptors is recognized by the memory even when the descripter information it is presented with happens to be incomplete or sketchy. Recognition of an input is manifested by the neural net converging and settling into one of its nominal states corresponding to that stored entity which most resembles the input. It is clear that this operation by itself may not be useful unless some means for identifying or recognizing the outcome of the nearest neighbor search is incorporated. This suggests, in the absence of such "recognizer", that auto-associative memories may be useful as building blocks of more complex processors in which higher order hierarchical processing takes place, in a manner not yet fully understood. Such processing would have the aim of first reducing the information content of the signals flowing concurrently through such a network of associative CAMs and then fusing their outcomes into a single meaningful concept or perception. In biological systems, such perception is thought to be associated with a prescribed trace of neural activity or spatio-temporal firing pattern in the cortex. Associative memory modules by themselves become more meaningful when used in a hetero-associative mode where the

learning set of descriptors or classifiers is not stored by auto-association but by hetero-association with reference entities that are more easy to recognize i.e., reference entities that are compatible with the "recognizer" such as word labels or pictorial representations in the case of a human observer or coded analog or digital outputs suitable for activating motor functions in robotic and artificial intelligence systems. Hetero-associative storage and recall of partial information can play an important role in smart sensing and automated recognition systems as was recently demonstrated with an example of identifying different types of aircraft from partial information about their sinogram classifiers [42] and individual faces from edge enhanced versions as classifiers [43].

3. CONCLUDING REMARKS

At this stage of their early development one can only venture to say that neural net models offer a new and intellectually stimulating approach to computing and information processing. The approach "dove-tails" very well the capabilities of optics and compliments them. While optics naturally provides the parallelism and massive interconnectivity needed in the implementation of neural nets and their models, these on their part provide the robustness, fault tolerance, and power of nonlinear processing and feedback that are generally lacking in optical processing. The combination results in systems that can no longer be characterized by the usual measures of convolution and impulse response (variant or invariant) because superposition no longer applies but may have to be characterized instead by stable states or modes in the N-dimensional phase-space of an N-neuron network. Optical analogs of even highly simplified models of neural nets exhibit a high degree of robustness and fault-tolerance, and can be

implemented as content addressable associative memories for use in computers and in smart sensing and automated recognition systems, or as networks for rapid collective solution of computationally complex tasks encountered in optimization, vision, and inverse problems. They may also provide a tool for the study of nonlinear dynamics, chaos, and even clinical studies of mental disorder.

4. REFERENCES

1. Widrow, B. and M.E. Hoff, "Adaptive Switching Circuits", WESCON Convention Record, Part IV, pp. 96-104, 1960. Also in Self-Organizing Systems, M.C. Yovitz, et. al. (Eds.) Spartan 1962.

2. Nakano, K., "Association-A Model of Associative Memory", IEEE Trans. on Systems Man & Cybernetics., Vol. SMC-2, pp. 380-388, 1972.

3. Kohonen, T., "Correlation Matrix Memories", IEEE Trans. on Computers, Vol. C-21, pp. 353-359, (1972).

4. Kohonen, T., Associative Memory, Springer Verlag, Heidelberg (1978). Also "Self-Organization and Associative Memory", Springer Verlag, (1984).

5. Willshow, D.J., "A Simple Network Capable of Inductive Generalization", Proc. Roy. Soc. London, Vol. 182, pp. 233-247, 1972.

6. Anderson, J.A., et. al., "Distinctive Features, Categorical Perception, and Probability Learning: Some Applictions of a Neural Model", Physilogical Review, Vol. 34, pp. 413-451, 1977.

7. Hopfield, J.J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Proc. Natl. Acad. Sci., USA, Vol. 79, pp. 2554-2558, 1982. Also, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-State Neurons", Proc. Natl. Acad. Sci., USA, Vol. 81, 1984.

8. Grossberg, S., Studies of Mind and Brain, R. Reidel Pub. Co., Boston, 1982.

9. Sanderson, A.C., and Y.Y. Zeevi, (Eds.), Special Issue on Neural and Sensory Information Processing, IEEE Trans. on Syst. Man and Cybernetics, Vol. SMC-13, 1983.

10. Tikhonov, A.N. and V.Y. Arsenin, "Solutions of Ill-Posed Problems", Winston and Sons, Washington, D.C. 1977.

11. Poggio, T. and C. Koch, "Ill-Posed Problems in Early Vision: From Computational Theory to Analog Networks", Proc. Roy. Soc. London, Vol. B226, pp. 303-323, 1985.

12. Farhat, N. and S. Miyahara, "Super-resolution and Signal Recovery Using Models of Neural Networks", Technical Digest, Spring 86 OSA Topical Meeting on Signal Recovery and Synthesis II, Honolulu, Hawaii, pp. 120-123, 1986.

13. Hopfield, J.J. and D.W. Tank, "Neural Computations of Decisions in Optimization Problems", Biol. Cybernetics, Vol. 52, pp. 141-152, 1985.

14. Ross, W., D. Psaltis, and R. Anderson, "Two-Dimensional Magneto-Optic Spatial Light Modulator for Signal Processing", Optical Engineering, Vol. 22, pp. 485-490, 1983.

15. Porada, Z., "Thin Film Light Amplifier with Optical Feedback", Thin Solid Films, Electronics and Optics, Vol. 109, pp. 213-216, 1983.

16. Gibbs, H.M., and N. Peyghambrian, "Nonlinear Optical Mechanisms and Devices for Optical Computing", Technical Digest, OSA Topical Meeting on Optical Computing, Incline Village, Nev. pp. MC1-1 o MC1-3.

17. Psaltis, D. and N. Farhat, "A New Approach to Optical Information Processing Based on the Hopfield Model", Digest of the 13-th Congress of Intern. Commiss. on Optics, ICO-13, Saporro, Japan, 1984. Also, "Optical Information Processing Based on Associative-Memory Model of Neural Nets with Thresholding and Feedback", Opt. Lett., Vol. 10, pp. 98-100, 1985.

18. Farhat, N.H., et. al., "Optical Implementation of the Hopfield Model", App. Optics, Vol. 24, pp. 1469-1475, 1985.

19. Farhat, N.H., and D. Psaltis, "Architectures for Optical Implementation of 2-D Content Addressable Memories", Digest OSA Annual Meeting, Wash., D.C., p. 58, 1985.

20. Lee, K.S. and N.H. Farhat, "Content Addressable Memory with Smooth Transition and Adaptive Thresholding", Digest OSA Annual Meeting, Wash., D.C., p. 48, 1985.

21. Psaltis, D., et. al., "Acousto-Optic Implementation of Neural Network Models", Digest OSA Annual Meeting, Wash. D.C., p. 58, 1985.

22. Condon, D.J., et. al., "Optical Window Addressable Memory", OSA Topical Meeting on Optical Computing, Incline Village, Nov. 1985, (Post deadline paper).

23. Fisher, A.D., C.L. Giles, and J.N. Lee, "Associative Processor Architectures for Optical Computing", J. Opt. Soc. Am., A, Vol. 1, p. 1337, 1984. Also, "An Adaptive, Associative Optical Computing Element", Technical Digest, OSA Topical Meeting on Optical Computing, Incline Village, Nev., pp. WB4-1 to WB4-4, 1985.

24. Fisher, A.D., and C.L. Giles, "Optical Adaptive Associative Computer Architectures", Proc. of the IEEE COPCOM Spring Meeting, IEEE Computer Society, IEEE Cat. No. CH2135-2/85/0000/0342, pp. 342-334, 1985.

25. Fisher, A.D., et. al., "Implementation of Adaptive Associative Optical Computing Elements", SPIE, Vol. 625, Bellingham, WA., 1986.

26. Sofer, B., et. al., "Associative Holographic Memory with Feedback, Using Phase-Conjugate Mirrors", Opt. Lett., Vol. 11, pp. 118-120, 1986.

27. Anderson, D.Z., "Coherent Optical Eigenstate Memory", Opt. Lett. Vol. 11, pp. 56-58, 1986.

28. Yariv, A. and S.K. Kwong, "Associative Memories Based on Message-Bearing Optical Modes in Phase-Conjugate Resonator", Opt. Lett., Vol. 11, pp. 186-183, 1986.

29. Psaltis, D. et. al., "Shift Invariance in Optical Associative Memories", SPIE, Vol. 625, Bellinghm, WA., 1986.

30. Cohen, M., "Distributed Computation in Neural Networks and Their Optical Analogs", OSA Topical Meeting on Optical Computing, Incline Village, Nev., 1985. (Post deadline paper).

31. Mirsalehi, M.M. and T.K. Gaylord, "Multi-Level Coded Residue-Based Content-Addressable-Memory Optical Computing", Technical Digest, OSA Topical Meeting on Optical Computing, Incline Village, Nev., pp. WB4-1 to WB1-4, (1985).

32. Abu-Mustafa, Y.S. and J.M. St. Jacques, "Information Capacity of the Hopfield Model", IEEE Trans. on Inf. Theo., Vol. IT-31, pp. 461-464, 1985.

33. McElice, R.J., et. al., "Capacity of the Hopfield Associative Memory", Caltech Report, 1986.

34. Venkatesh, S. and D. Psaltis, "Information Storage and Retrieval in Two Associative Nets", Submitted to IEEE Trans. on Info. Theo.

35. Von Neuman, J., Probabilistic Logics and Synthesis of Reliable Organism From Unreliable Components in Automata Studies, C.E. Shannon and J. McCarthy (Eds.), Princeton Univ. Press, Princeton, N.J., pp. 48-98, 1956.

36. Fukushima, K., "Neocognition:A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", Biol., Cybern., Vol. 36, pp. 193-202, 1980.

37. Hubble, D.H., "The Visual Cortex of the Brain",Scientific American, Vol. 209, pp. 54-63, 1963.

38. Hubble, D.H., and T.W. Weisel, "Functional Architecture of Macaque Monkey Visual Cortex", Proc. Roy. Soc. Lond. B., Vol. 198, pp. 1-59 1977.

39. Tootel, R.B., et. al., "Spatial Frequency Columns in Primary Visual Cortex", Science, Vol. 214, pp. 813-815, 1981.

40. Orban, G.A., Neuronal Operations in the Visual Cortex, Springer-Verlag, New York, see for example, sections 7.3.4 and 15.1.3., 1984.

41. Jones, J.P., and L.A. Palmer, "Simple Receptive Fields of Cat Striate Cortex: A Comparison with Gabor Functions in Two Dimensions of Space and Two Dimensions of Spatial Frequency", University of Pennsylvania - private communication, 1985.

42. Farhat, N. et. al., Optical Implementation of 2-D Neural Nets and Their Application in Recognition of Radar Target", Presented at the Neural Networks for Computing Conference, Snowbird, Utah, 1986.

43. Psaltis, D., et. al., "Nonlinear Discriminant Functions and Associative Memories", Presented at the Neural Networks for Computing Conference, Snowbird, Utah, 1986.

SCHEME FOR ENHANCING THE FRAME RATE OF MAGNETO-OPTIC SPATIAL LIGHT MODULATORS

N. H. Farhat Z. Y. Shae

University of Pennsylvania
Electrical Engineering Department
Phila. PA 19104

Abstract

High speed computer-controlled spatial light modulators play an important role in optical signal processing. Of interest is a commercially available nonvolatile Litton/Semetex magneto-optical spatial light modulator (MOSLM) which is limited in achievable frame rate by thermal dissipation and possible magnetic forces. Methodologies for improving the frame rate of the MOSLM are investigated. The findings described here show that a parallel-serial addressing scheme aided by a strong externally applied magnetic field bias could be employed to achieve high frame rates. A driver circuit capable of exercising the 48 $\times$ 48 MOSLM at a frame switching time of better than 1/1000 second is described and its performance evaluated.

1 Introduction

High speed programmable optical masks play an important role in optical signal processing. Such programmable optical masks are generally referred to as Spatial Light Modulators (SLM). One commercially available nonvolatile SLM is a magneto-optic device marketed either by Litton [1-3] under the trade name LIGHT-MOD or by Semetex Corp. under the trade name SIGHT-MOD [4-5].

In this paper we use the term MOSLM to describe these devices because of the magneto-optic effect utilized in this operation. The device has binary pixel transmittance and commercial units can be readily raster driven at a rate of few tens of frames per second. In exploring operation at higher frame rates parallel addressing of rows (or columns), one at a time, is an obvious approach. Higher frame rates are in principle possible in these devices because of the known extremely short pixel switching time ($\sim 10^{-9}$ sec [1]). In order to explore the speed at which this device can operate, we built a high speed interface circuit, not offered by the manufacture, for this device. We encountered two device limitations that can be destructive unless special provisions are made to circumvent them. One is the inability to dissipate heat produced by ohmic losses in the current bearing addressing wires specially when parallel addressing is attempted; the other is possible mechanical magnetic forces that can fracture the device. These are caused by the interaction between the current bearing addressing wires and the magnetic field established by the external drive coil of the device. When too many addressing wires are active simultaneously as would be the case in a parallel addressing scheme, the large net current will make it very difficult to realize the entire driver circuit in a single VLSI chip. Moreover, we found that one reaches the ohmic losses limit of the device (no cooling) before a parallel addressing scheme can show its very high frame rate benefit. These considerations showed us quickly the impracticality of a fully parallel addressing approach to operation at high frame rate. Therefore, we developed a "parallel-serial addressing" scheme in which the driver circuit buffers all the output data bits from a computer controller's parallel digital output interface card. Every bit corresponds to a desired state of a pixel in the MOSLM. The driver circuit then address in parallel a small number of pixels at a time, and process them at a very high speed. In other word, The driver circuit staggers the addressing time instant of data bits from the computer's parallel digital output interface card. By doing this, we overcome the limitation of magnetic mechanical forces without seriously degrading the speed. We also reduce the magnitude of the required current by using a strong auxiliary external magnetic field bias. As a result ohmic losses are reduced considerably because of the lower drive current pulses. This relatively low addressing current feature suggests that building the driver circuit (except the high current and possibly

high voltage driving components for the external coil) in a customer's VLSI chip is possible. A driver circuit capable of exercising the 48 $\times$ 48 MOSLM at approximately 1000 frames per second without any special cooling means has been built successfully and is described below. The work shows that the switching threshold of a MOSLM plays an important role in determining the achievable frame rate.

The operational principle of the MOSLM is described in section 2, and its possible limitations are discussed in section 3. The parallel-serial addressing scheme is then introduced in section 4, and the methods used to relax ohmic losses are described in section 5. Experimental results are presented in section 6.

2 Operation Principle of A MOSLM

2.1 Magneto-Optic Effect — The Faraday Rotation

The Faraday rotation is a property of transparent substances that causes a rotation of the plane of polarization of light traversing the substance when the material is subjected to a magnetic field. The rotation is proportional to the component of the magnetic field along the direction of propagation of the light. It can be shown [6] that when linearly polarized light propagates along the optical axis of a uniaxial material with Faraday rotation, the material will support right and left circular polarized wave modes of equal magnitude. Further, the indices of refraction associated with these two modes are different. As a result, the output light will be linearly polarized with a different polarization direction than that of the input light. The angle between the directions of the input and output polarization is called the Faraday rotation angle. This angle is dependent on the Verdet constant [6] of the material and is proportional to the thickness of the material and the magnitude of the applied magnetic field.

2.2 Function of A Switching Element

The magneto-optic switching element or pixel in a MOSLM consists of a bismuth substituted transparent iron garnet film grown on a nonmagnetic single crystal (gadolinium galium garnet crystal) substrate. It is a uniaxial crystal

with its uniaxial axis oriented perpendicular to the surface plane of the film. The switching element is practically a bistable magnetic domain with an internal magnetic field always parallel to the uniaxial axis. The two stable directions of its internal magnetization can be electrically switched rapidly. Once we switch the direction of the internal magnetization of the switching element, we switch the direction of the Faraday rotation angle from clockwise to counterclockwise or vice versa. By sandwiching the element between a polarizer at the input face and analyzer at the output, transmitted monochromatic light intensity I_t can be controlled according to the formula [7],

$$I_t(\pm\theta_F) = I_o\beta \exp(-\alpha t)[\sin^2(\phi_{pa} \pm \theta_F)] \tag{1}$$

where I_o is the incident light intensity, α is the optical absorption constant of the material, β is the loss due to the polarizer and analyzer, t is the material thickness, ϕ_{pa} is the relative angle between the polarizer and analyzer axes measured from the extinction position (crossed out position), and θ_F is the magnitude of Faraday rotation.

The difference in the transmitted light intensity of pixels in the two domain states can be used in display and optical computation.

2.3 The Structure of A MOSLM

The MOSLM is a rectangular array of magneto-optic switching elements. Coincident orthogonal current drive-lines are produced by conductor deposition on both sides of the iron garnet film. The magnetic field generated by current flowing in a single conductor is designed to be insufficient to switch the state of a single pixel. The X-Y coincident current drive-lines make it possible to switch individual pixels by random selection. The combined magnetic fields will switch the state of selected element only. A scanning electronic microscope (SEM) picture of a MOSLM is shown in Fig. 1. It shows clearly that each switching element is addressed by two orthogonal wires. Each wire deforms into a small loop at the corner of each switching element where the alteration of state of a pixel is initiated. When an element, at the intersection of the activated X-Y conductors, is addressed to reverse its state, a magnetic domain is formed in this corner and will be enlarged by the propagation of the domain

wall through the thickness of the film. When the domain wall reaches the bottom of the film, the element has been nucleated. Removal of the drive current at that time would result in the element being demagnetized. By maintaining drive currents until the domain wall has propagated to the opposite corner, the element will complete switching and will be stable i.e. remain in that state until the next nucleating process occurs. We also observe from Fig. 1 that some addressing wires fail to form a proper loop in some elements because of imperfect fabrication which result in the element not being able to switch its state in a normal addressing operation. The appearance of such imperfect elements can be seen in the display of Fig. 10. The above selective switching of the state of magnetization of pixels takes place in the presence of an external magnetic field produced by an external coil used to generate a magnetic field for assisting the element switching.

3 Possible Limitations

3.1 Ohmic Losses

The typical resistance of a drive-line in the device is about 30 ohms. The minimal current required to switch the element is 100 mA when a 500 mA current is used to drive the external coil to produce the external magnetic field bias (30 gauss) recommended in the device specifications. Accordingly, the thermal power due to ohmic losses generated by the current in one drive-line is 300 mW. The switching of one element with coincident currents in the two orthogonal drive-lines intersecting at the element would dissipate 600 mW in the MOSLM. Because the area of the device is so small (less than a half cm sq.),it is difficult to dissipate this much heat even when current pulses at duty cycle less than the duty cycle of unity assumed in the above consideration is utilized. Based on the manufacturer's information, the maximum estimated power dissipation that the device (uncooled) can sustain is 100 mW [8]. Accordingly, in order to operate the device safely, the current pulses with duty cycle less than 1/6 have to be used even when a conventional raster switching scheme (one element a time) is adopted. In addition to this ohmic losses problem, the effect of localization of the heat in the device should be taken into account. Because the heat conduc-

tivity of the MOSLM is not very good, the heat generated can be localized. In general, the temperature in the center part of a MOSLM will be higher than that in the outer parts and so would temperature near the addressing wires be higher than at distal parts. The temperature unbalance generates stresses that can easily fracture the crystal substrate. We happened to observe these ohmic loss effects by applying current in 17 drive-lines simultaneously to partially address one row of pixels in the device (One line is parallel to the X axis and 16 lines are parallel to the Y axis). Every two adjacent lines form one pair, there are 8 pairs along the Y axis. The current in each pair flowed mistakenly in opposite directions (should flow in the same direction in correct operation). Based on the magnetic force analysis presented later, this addressing arrangement happens to distribute magnetic force uniformly over the device, and the forces nearly canceled each other. Thus this parallel addressing arrangement rules out possible damage due to magnetic forces. The current magnitude in each wire was 300 mA, the pulse width was 1 μs, and the duty cycle was 0.5. The result was that the MOSLM shattered into pieces after approximately 3 seconds of operation. A photograph of the damaged device is shown in Fig. 2. We also show a photograph of the polarizer damaged by heat emitted from the MOSLM during this incident which was situated a distance of 1 cm from the polarizer.

3.2 Magnetic Force

3.2.1 Magnetic Force Between Drive-Lines

Based on magnetostatic field theory, parallel current carrying wires experience a magnetic force between them. The magnitude of the force is proportional to the magnitude of the current, the number of wires simultaneously carrying current and the relative distances between them. Assume there are 2 parallel wires each carrying current I with the distance between these two wires being d, and the length of each wire being infinite. The force per unit length between such two wires is given by [9]

$$F = \frac{\mu_o I^2}{2\pi d} \qquad [\frac{newton}{m}] \tag{2}$$

where $\mu_o = 4\pi \times 10^{-7} H/m$ is the permeability of free space. The force is attractive if the currents flow in the same direction, and repulsive otherwise. Assume we have N parallel wires numbered 1 to N each carrying the current I in the same direction. The force on wire 1 points to the center wire with magnitude F_1, and can be written as

$$F_1 = F\ell \sum_{i=1}^{N-1} \frac{1}{i} \tag{3}$$

where ℓ is the length of each wire. This series diverges if N is infinite [10]. This implies a significant force for large N. Similarly the force on wire 2 also points to the center wire with magnitude F_2, and can be written as

$$F_2 = F\ell \sum_{i=2}^{N-2} \frac{1}{i} \tag{4}$$

As a result, a general expression can be derived for the force exerted on the n-th wire,

$$F_n = F\ell \sum_{i=n}^{N-n} \frac{1}{i} \tag{5}$$

where n is less than or equal to $N/2$, and $\vec{F_n}$ = - $\vec{F_{N+1-n}}$. The forces always point to the center wire. The total force $\vec{F_A}$ in one half of the MOSLM is the sum of the forces of individual wires from 1 to $N/2$. For a numerical instance, when N=16, ℓ =0.5 cm, d=228 μm, and I=300 mA, we find $\vec{F_A}$ is on the order of 10^{-5} *newton*.

3.2.2 Magnetic Force Between The Bias Magnetic Field And The Drive-Lines

The external axial magnetic field will exert force on the perpendicular current carrying wire. This force is also proportional to the magnitude of the current drive wires and the number of drive wires activated simultaneously. Assume the external coil has length L, the number of turns per unit length is M, and the mean radius is a. The magnetic field at the end of this solenoid can be decomposed into two orthogonal components [11],

$$B_z = \frac{\mu_o MI}{2L} \tag{6}$$

$$B_r = \frac{\mu_o MI}{4L}\frac{r}{a} \tag{7}$$

where $\vec{B_z}$ is the magnetic field component in the axial direction of the solenoid, $\vec{B_r}$ is the radial or transversal magnetic field component, and r is the radial distance from the central axis. Assume there are N parallel wires each carrying current I, and these wires are uniformly spaced over the MOSLM extent. First we consider the force due to the component $\vec{B_z}$. The magnetic force on a single wire can be written as $\vec{F_B} = \ell\vec{I} \times \vec{B_z}$. The magnetic forces on each wire due to $\vec{B_z}$ have the same magnitude, and are all in the same direction. For a numerical instance, assume N=16, ℓ =0.5 cm, a=1 cm, L=1 cm, and I=0.5 A. The resulting force on each wire would be 3.2×10^{-4} *newton*. Therefore, the total force due to $\vec{B_z}$ acting on the 16 wires will be 3.4×10^{-3} *newton*. Next, we consider the force due to the component $\vec{B_r}$. Assume the N wires are equally distributed along the two side of the axis of the solenoid. Then $\vec{F_n}$ = - $\vec{F_{N+1-n}}$, where $\vec{F_n}$ is the force on the n-th wire. Similar analysis shows the total force $\vec{F_C}$ due to $\vec{B_r}$ is approximately 4×10^{-4} *newton*.

3.2.3 Total Magnetic Force

By superimposing the three magnetic force components described above we obtain the total force diagram shown in Fig. 3a , where the notation ⊗ designates current flowing into the plane of the paper. These force components are proportional to the number of current carrying wires, and the magnitude of the current in each wire. $\vec{F_A}$ is also inversely proportional to the distance between the wires. It is worth nothing that all these forces will be locally distributed and will nearly cancel each other if the currents in adjacent pairs of wires flow incorrectly in opposite directions as in the case of section 3.1. Accordingly, we preclude the magnetic force mechanism in the case of section 3.1. However, in the case of correct parallel addressing operation, all currents must flow in the same direction. Therefore, there will be a fracture line running along Y axis at the center of the MOSLM because of these magnetic forces if we exercise this device with a parallel addressing scheme. We happen to observe this effect when we attempted to repeat the parallel addressing experiment mentioned in section 3.1 (currents now flow correctly in the same direction) with the device air cooled and operating at a very low duty cycle of current pulse to avoid thermal damage. Again, 17 drive wires were activated simultaneously; one along the X axis, and 16 along the Y axis which are equally distributed over the width of the

MOSLM. The magnitude of current is 300 mA in each wire, the current pulse width is 1 μs, and the duty cycle is about 0.0001. The device split into 2 pieces after approximately one second of operation. The split occurred at a central directed line on the device as predicted by the magnetic force considerations as illustrated in pictorial view of Fig. 3.

4 The Parallel-Serial Addressing Scheme

The purpose of this scheme is to minimize the magnetic force the device has to sustain, while maximizing the operating speed. We give an example to illustrate the parallel-serial addressing scheme. The MASSCOMP 5400 computer has a parallel digital output interface card which can output 16 bits of data simultaneously at every 2 μs. If the driver circuit can process the 16 bits of data within 2 μs bit by bit, the system will have the same speed as that of direct parallel addressing scheme by the available computer. Using this parallel-serial or systolic scheme, we address a single pixel at a time and at a very high speed. we can avoid the problem of magnetic mechanical forces by staggering the switching time instant of the elements and reducing the number of wires that carry current simultaneously. This will, in turn, reduce the total instantaneous current required in the driver circuit and reduce the huge current spikes in the driver circuit. This relatively low addressing current feature will also make building the entire driver circuit into a single customer's VLSI chip practical. These are the fundamental concepts behind the parallel-serial addressing scheme. The success of this scheme will depend entirely on how fast each pixel of the MOSLM can be switched. Our experimental results presented later show that the switching time is on the order of 100 ns. Such a high pixel switching speed makes our system capable of processing all 16 bits of data within 2 μs i.e. an addressing serial bit rate of 8×10^6 bits/sec. Accordingly, if we use the MASSCOMP 5400 computer to provide the data pattern to the driver circuit, there will be no speed difference between the parallel-serial addressing scheme and parallel addressing scheme. Moreover, if we have a computer which can output 32 bits of data at a time once every 2 μs, the driver circuit of the parallel-serial addressing scheme can process the data 2 bits at a time, i.e. by driving two at a time, addressing every 100 ns. In this manner, the driver

circuit completes the processing of the 32 bits of data within 2 μs. Again, the parallel-serial addressing scheme can virtually have the same speed as that obtained had we been able to address the device directly in parallel in batches of 32 pixels at a time. Generally speaking, the parallel-serial addressing scheme provides a flexible way to drive the MOSLM with two benefits: one is that the device will sustain minimal magnetic force while operating at the maximum system speed; the other is that the total instantaneous maximum driver current required will be the minimized. However, the design of the driver circuit employing a parallel-serial addressing scheme is much more complicated than that employing a parallel addressing scheme. In order to get the same MOSLM frame rate as that offered by the parallel addressing scheme in the present system, the clock rate of the driver circuit employing a parallel-serial addressing scheme must be 16 times faster than that of a parallel addressing scheme. To demonstrate these ideals, we have designed and built a driver circuit employing the parallel-serial addressing scheme. A block diagram of which is shown in Fig. 4.

5 Schemes for Relaxing Ohmic Losses

As stated earlier the thermal power generated in the device by ohmic losses produced by currents in addressing wires will limit the operating speed of the MOSLM because of the finite power dissipation capability of the uncooled device. There are several ways to deal with this problem. The first approach is either to reduce the resistance of the addressing wires by increasing their thickness, or to reduce the switching threshold of the magnetic domain (pixels) of the device through ion implantation or selection of improved magneto-optic material [12]. The drive-line current required for switching decreases with decreasing the switching threshold. The second approach to this problem is the setting up of a good cooling system. The active area of the MOSLM is however too small for effective air cooling at a high duty cycle of current pulses. Since the device is used ordinary in a transmission mode, the use of any cooling substrates mean they have to be transparent in the operating wavelength of the MOSLM, good conductors of heat, and do not interact between the MOSLM crystal and the cooling material. If the device operates in the reflection mode,

it needs cooling materials which are optically flat, have good reflectivity, are good conductors of heat, and are without side effects. Such cooling schemes obviously complicate the system. The third approach is to make use of the thermal properties of the switching element material. The switching threshold is inversely proportional to the operation temperature before the curie temperature is reached [13,14]. However, this approach may have practical difficulties due to the fact that the temperature distribution over the device is not uniform and the temperature of each element is varying with time. In order to guarantee perfect switching operation for each element, a very intelligent driver circuit would be required to monitor the temperature of all pixels so that a suitable amplitude and width of the addressing current pulse can be determined. The fourth approach in dealing with operating speed limitation is to reduce the magnitude of the current needed to switch pixels by using a strong external magnetic field bias in addition to that produced by the magnetic coil. Note that using a strong external magnetic field bias to reduce the ohmic losses without using the parallel-serial addressing scheme might increase the hazard of damage by magnetic forces. However, using the parallel-serial addressing scheme alone without using a strong external magnetic field bias can not overcome the ohmic losses. Therefore, these two schemes have to work together to achieve high frame rates. Erasure of all pixels at once can be done by applying a strong external magnetic field only. There is no current required in the addressing wires in this erasing operation. Then, only write operations require currents flowing in the addressing wires. Accordingly, this scheme will reduce the heat dissipation in the MOSLM by a factor of one half. We have chosen to adopt and investigate the last two methodologies.

6 Experimental Results

6.1 Basic MOSLM Characteristic Measurements

A series of experiments to study the magnetic switching properties of a MOSLM were carried out. A schematic diagram of the arrangement used is shown in Fig. 5 when a polarized laser is used as the illumination source. A pictorial view of the setup is shown in Fig. 6 with the laser light source replaced by

an incoherent light source plus a polarizer. The device can thus be operated with coherent or incoherent light. Shown in figure 6 is the hollow cylindrical (ring type) magnet used to produce the strong auxiliary external magnetic field bias. The axial magnetic field distribution of this magnet (model 008-01-8 manufactured by Avco Everett Research Lab.) was measured using a gauss-meter (Radio Frequency Lab. model 1890). The measured axial magnetic distribution or magnetic field profile is shown in Fig. 7. The permanent magnet is moved close to the MOSLM until all pixels switch their states. We recorded the distance from the permanent magnet to the MOSLM crystal. From Fig. 7, this distance corresponds to a magnetic field strength which is called the switching threshold of a MOSLM's pixels. We observed that the switching threshold of pixels is not always the same even within the same MOSLM chip. For two available MOSLMs we recorded a switching threshold of 200 gauss and 300 gauss respectively. The threshold variation in these two devices is therefore large. This fact implies that the switching threshold is strongly influenced by the manufacturing process. Next a 500 mA current was applied in the external coil and we measure the magnetic field profile along the central axis of the coil. The measured field profile is shown in Fig. 8. Based on the data in Fig. 8, the 500 mA current in the external coil is equivalent to 30 gauss bias in the plane of the MOSLM. The inductance of the coil is 0.782 mH and its resistance is 1.26 ohm. The magnetic field produced in the plane of the MOSLM by the permanent magnet is easily adjusted by changing the axial distance between the magnet and the MOSLM. The use of the permanent magnet provides a convenient way of studying the relation between the external magnetic field bias and the required addressing wire current needed to switch pixels as will be discussed in the next section.

6.2 The Effect of A Strong External Magnetic Field Bias on The Required Drive-Line Current

Our aim was to determine the degree to which the switching drive current could be reduced when an external magnetic field bias is used. First, we used the MOSLM which has a switching threshold of 200 gauss. This device requires a 100 ns, 100mA coincident current pulses in the X and Y to write a pixel with

a 30 gauss bias from the external coil but without the help of the external permanent magnet. The writing current pulse is shown in Fig. 9a. We then place the permanent magnet 2.7 cm away from the MOSLM. This is a distance just beyond the critical distance where some of the pixels start to switch their states due to the strong magnetic field provided by the permanent magnet. This arrangement establishes a 180 gauss bias and reduces the current pulse required to write one pixel to 17 mA peak amplitude and 50 ns duration as shown in Fig. 9b. From Figs. 9a and 9b, we can see that the external magnetic field bias not only reduces the current magnitude but also the current pulse width needed for writing a pixel. Therefore, the external magnetic field bias has dramatic effect on the achievable frame rate of the MOSLM. We can switch all the pixels at once by applying reverse current of 5 A in the external coil when there is no permanent magnet in the setup. However, with the unidirectional strong bias provided by the permanent magnet, we need a larger current in the external coil to erase the pixels. This current, in turn, generates heat in the coil. This presents a hazard to the MOSLM crystal because of the coil's proximity to the crystal. The present driver circuit can not supply a current large enough to erase the pixels in this situation. Therefore, we place the permanent magnet 4.5 cm away from the MOSLM with a 500 mA current in the external coil. In this situation, the total bias is 110 gauss, of which 80 gauss are provided by the permanent magnet, and 30 gauss by the coil. With this arrangement, we can write the pixel with a current pulse of 70 ns duration and 66 mA peak amplitude, and erase all pixels simultaneously by applying 7 A reverse current in the external coil. The shape of the writing pulse is shown in Fig. 9c. Based on the 100 mW power dissipation constraint, we estimate that this particular device (with a 200 gauss switching threshold) can be exercised at approximately 400 frames per second without the permanent magnet bias, and at approximately 1000 frames per second with the permanent magnet bias. Similar considerations of the second MOSLM (with a 300 gauss switching threshold) yields the following results: with a 500 mA current in the external coil and no permanent magnet bias, the required writing current pulse is 200 mA peak and 250 ns duration and the erasing current is 7.5 A. With the permanent magnet 6.5 cm away (40 gauss bias) and a 550 mA current in the coil, the required writing current pulse is reduced to 170 mA peak and 180 ns duration and the erasing

current is 8 A. Based on the 100 mW power dissipation constraint, this particular MOSLM can achieve 60 frames per second without the permanent magnet bias, and 120 frames per second with the permanent magnet bias. The above results of these two devices show that switching threshold plays an important role in the achievable frame rate and should be used as a design parameter in future MOSLM fabrication. The benefit of using a permanent magnet to generate the strong external magnetic field bias and increase frame rate in the present scheme is apparent. Finally as an illustration of the capability of the drive circuit with the parallel-serial addressing scheme, one of the two MOSLMs (the 300 gauss switching threshold device) was addressed with a complete pattern (see Fig. 10) in about 0.6 msec. Missing portions of the pattern in Fig. 10 are due to disconnection of some addressing wires, imperfection in magnetic domain triggering loops of some pixels (see reference of Fig. 1 in section 2.3) or merely bad pixels in this particular device.

7 Conclusions

Possible limitations and methodologies for enhancing the frame rate of a MOSLM were investigated. The findings indicate that direct parallel addressing is impractical for present commercially available devices because of thermal and possible magnetic stresses and that the single parameter determining the highest frame rate is the switching threshold of pixels in the device. In our opinion, the parallel-serial addressing scheme aided with a strong external magnetic field bias provides the best strategy for achieving high frame rates with present commercially available MOSLM device. Note that the switching speed of the magnetic domain itself in these device can be very fast (few tens of ns). Thus the speed limitation is primarily due to the heat generated by the ohmic losses in the addressing wires. Further enhancement of the frame rate of MOSLMs should be possible by exploration of new materials and fabrication techniques that lead to uniform reduction of the switching threshold for all pixels and reduction of the resistance of the addressing wires.

8 Acknowledgement

This work was supported by grants from DARPA/NRL, the Army Research Office, and the University of Pennsylvania Laboratory for Research on the Structure of Matter (LRSM). The SLMs were generously furnished by Litton and Semetex corp.. The printed circuit boards for the parallel-serial addressing scheme driver were fabricated at Scientech Electronics corp. of Taiwan. Their supports are gratefully acknowledged. The authors also wish to thank Dr. Takeshi Egami of LRSM for kindly furnishing the SEM pictures of the MOSLM.

9 References

1. W. E. Ross, et. al., "Fundamental characteristics of the Litton iron garnet magneto-optic spatial light modulator," SPIE Vol. 388, pp. 1-10 (1983)

2. B. Hill, "Optical transmissivity of magneto-optic switching elements," Philips J. Res. Vol. 38, pp. 150-163 (1983)

3. W. E. Ross, et. al., "Two-dimensional magneto-optic spatial light modulator for signal processing," SPIE Vol. 341, pp. 191-198 (1982)

4. "Proposal for application of Litton Magneto-Optic Light Valve to optical correlation in target acquisition," Litton System Inc., Data System Division, 8000 Woodley Avenue, Van Nuys, California 91409

5. Semetex Iron Garnet H (magnetically) Triggered Magneto Optic Device, Description Brochure

6. Yariv, Yeh, "Optical waves in crystal," chapter 4, John Wiley and Sons (1984)

7. G. S. Almasi, "Magnetooptic bubble-domain devices," IEEE Tran. on Magnetics, pp. 370-373, Sep. (1971)

8. T. Maki, Litton System Inc., Data System Division, 8000 Woodley Avenue, Van Nuys, California 91409, Private Communication, (1986)

9. Lorain, Corson, "Electromagnetic wave and field," Freeman, San Francisco, (1971)

10. G. Arfken, "Mathematical methods for physicists," Academic Press Inc. (1985)

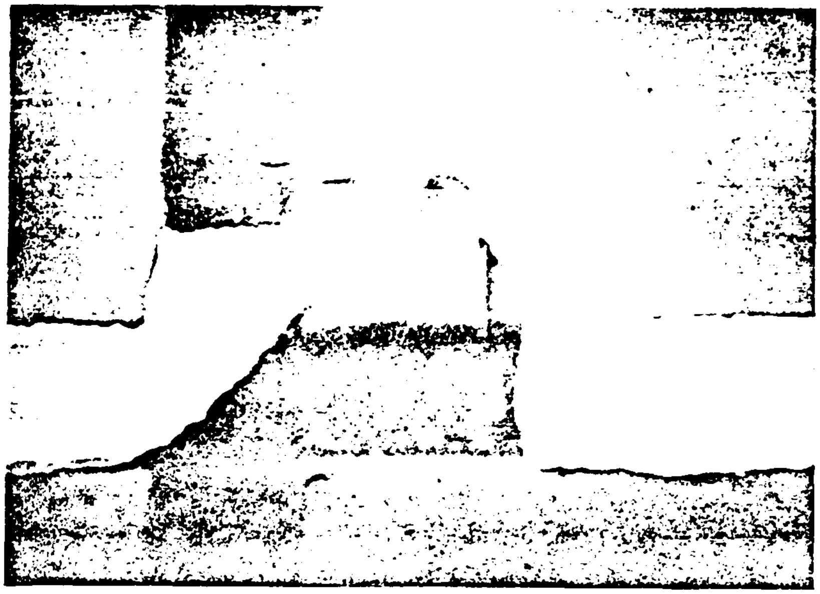
11. J. D. Jackson, "Classical electrodynamics," John Wiley and Sons, pp. 206 (1975)

12. G. R. Pulliam, et. al., "Large stable magnetic domains," J. Appl. Phys. Vol. 53(3), pp. 2754-2758, March (1982)

13. A. H. Eschenfelder, "Magnetic bubble technology," Springer-Verlog Berlin Heidelberg New York (1980)

14. J. P. Krumme, et.al., "Thermomagnetic recording in thin garnet layers," Appl. Phys. Lett., Vol. 20, No. 11, pp. 451-453 (1972)

1250 X



320 X

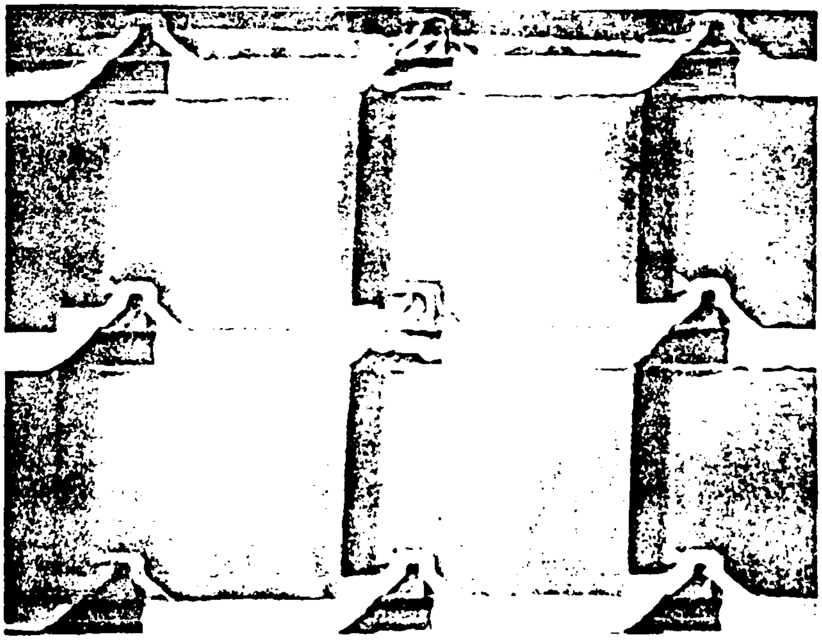


Fig. 1 SEM photograph of MOSLM. Showing the addressing wires forming a magnetic loops at the corner of each pixel where domain switching is initiated

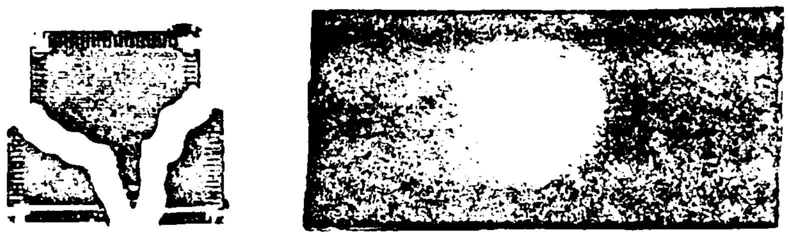


Fig. 2 (Left) Thermally damaged device and (right) polarizer damaged by heat generated in the MOSLM

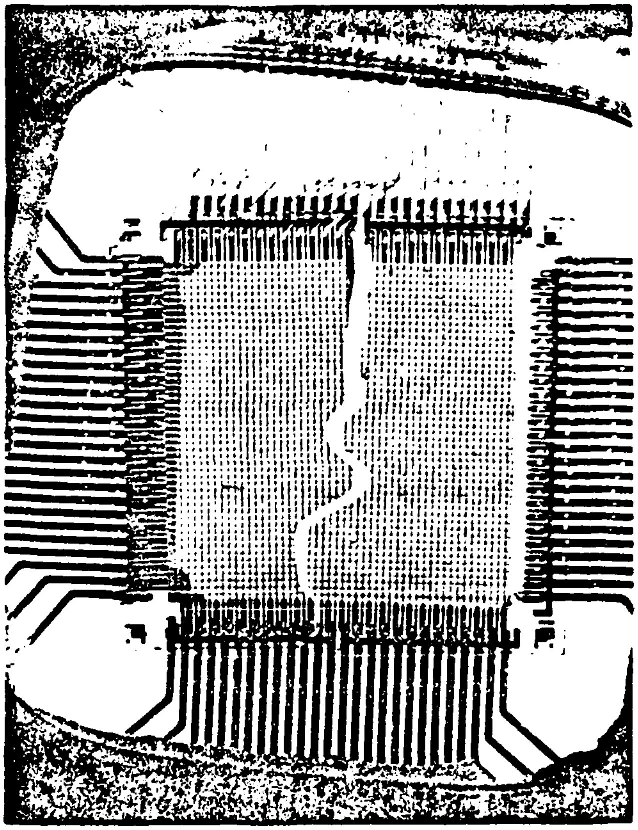
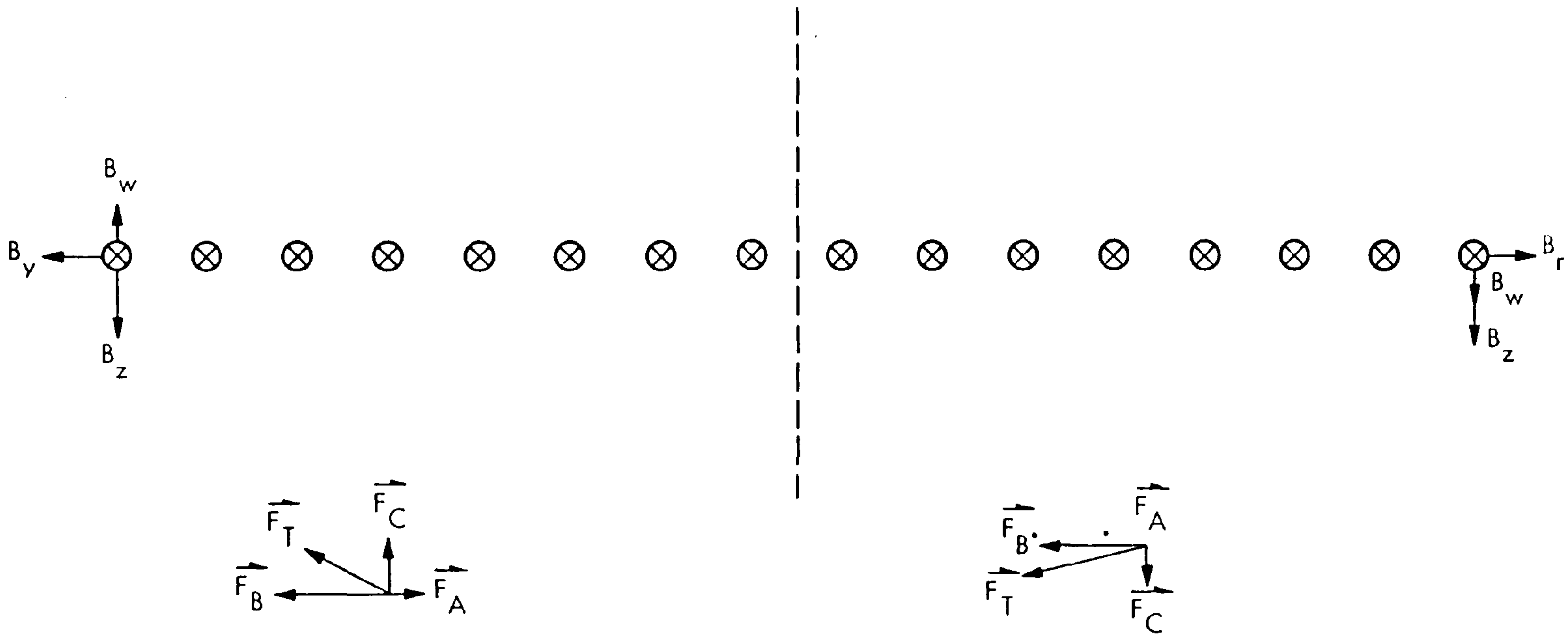


Fig. 3 Device damaged possibly by magnetic forces

Fig. 3a Magnetic Force Diagram in MOSLM



B_z IS AXIAL MAGNETIC FIELD COMPONENT OF SOLENOID, IT GENERATES FORCE COMPONENT F_B. B_r IS TRANSVERSAL MAGNETIC FIELD COMPONENT OF SOLENOID, IT GENERATES FORCE COMPONENT F_C. B_w IS THE SUMMATION OF MAGNETIC FIELD COMPONENT FROM WIRES, IT GENERATES FORCE COMPONENT F_A. F_T IS THE TOTAL MAGNETIC FORCE ON ONE HALF OF THE MOSLM

Fig. 4 Driver interface circuit block diagram employing the parallel-serial addressing scheme

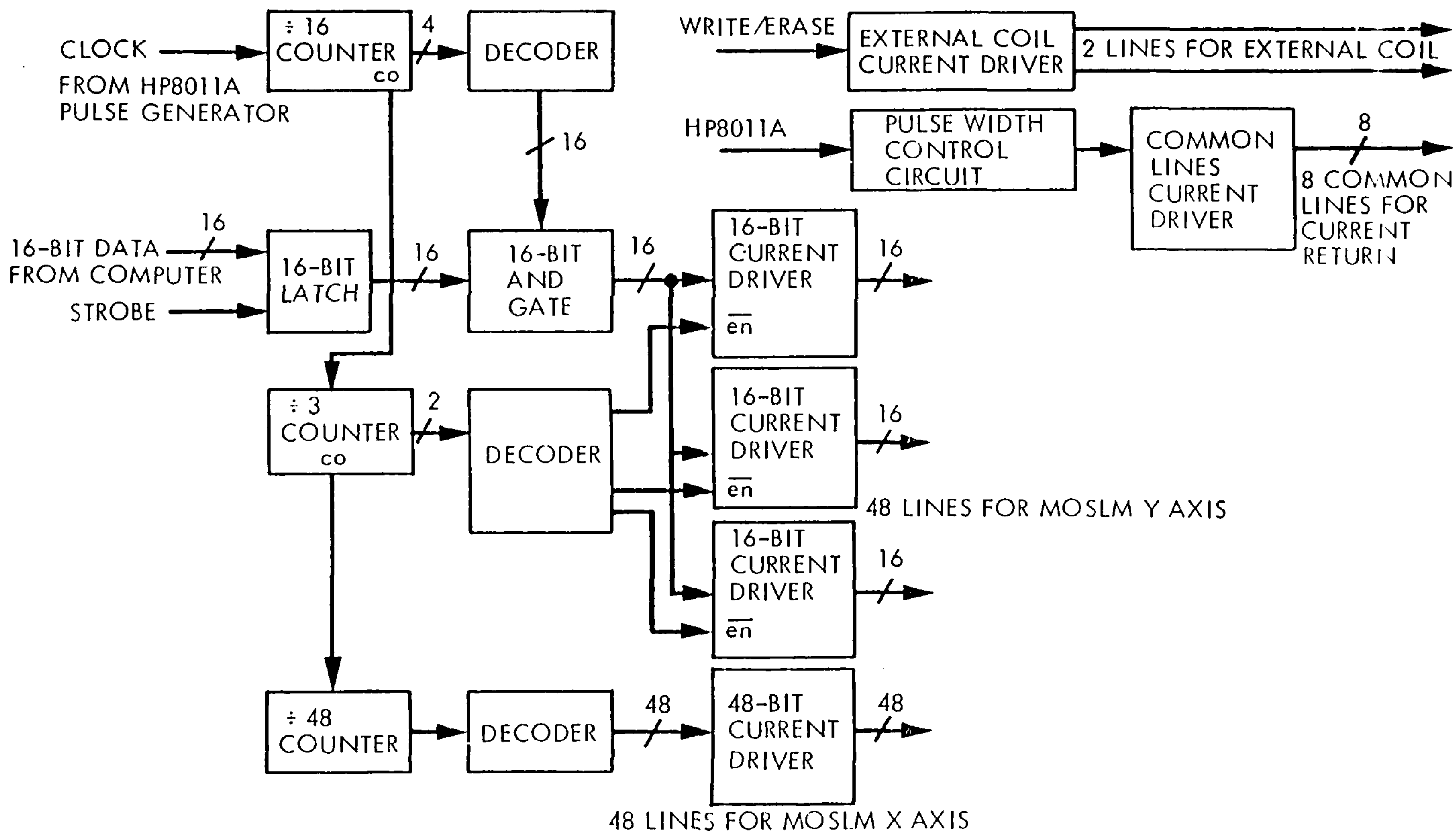
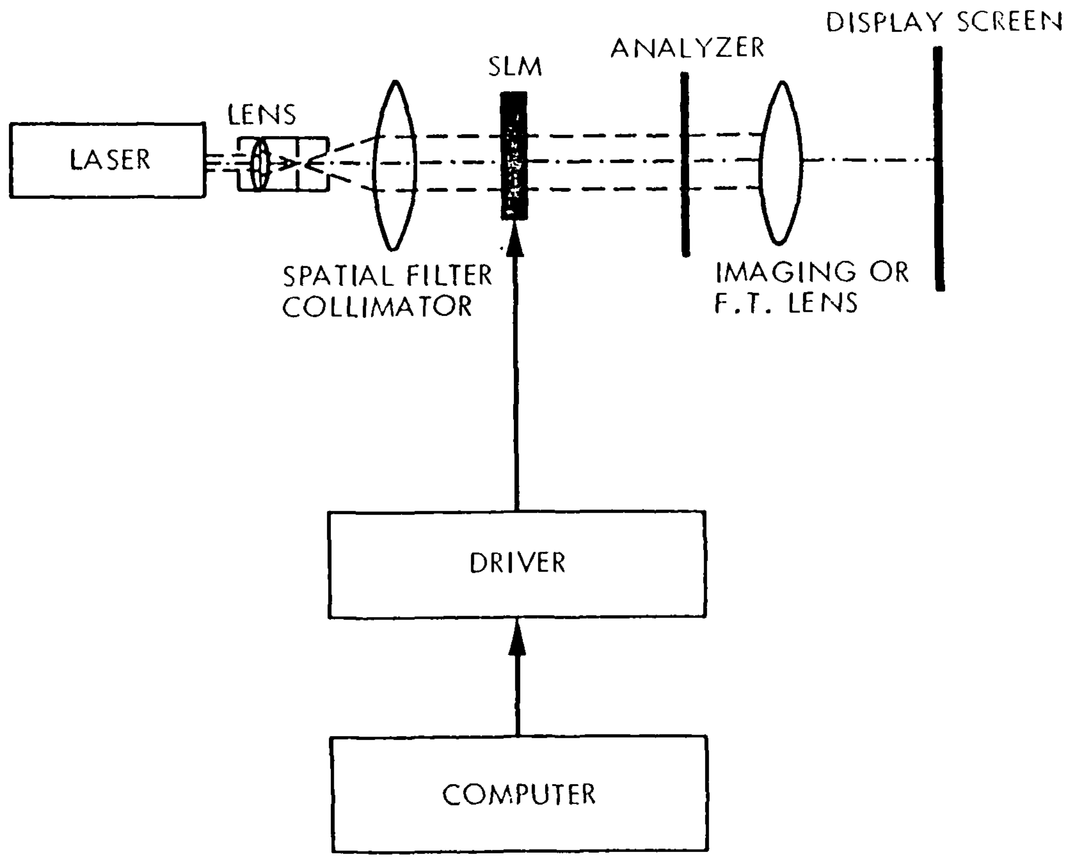


Fig. 5 System Schematic With Laser Light Source



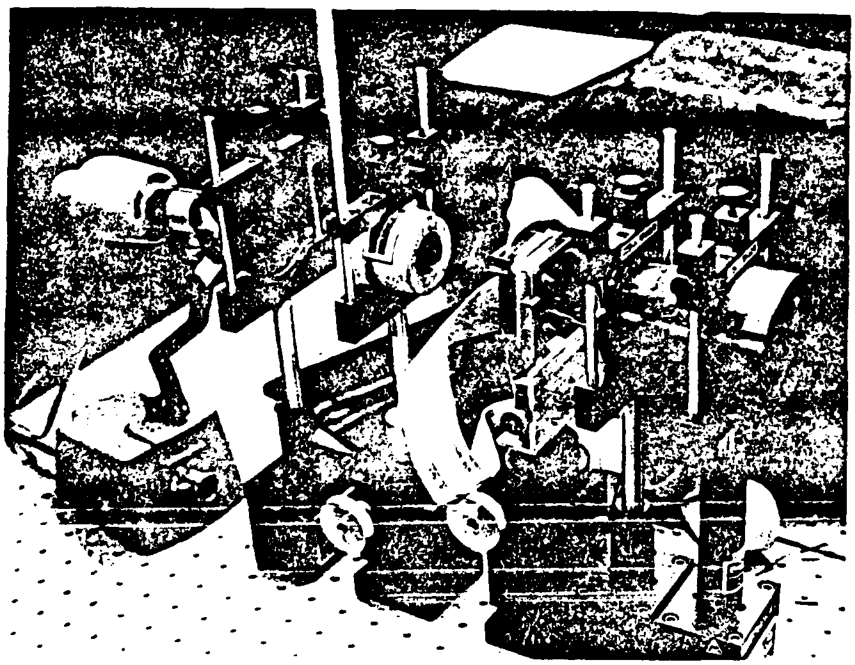


Fig. 6 Pictorial view of experimental arrangement with the incohent illumination source

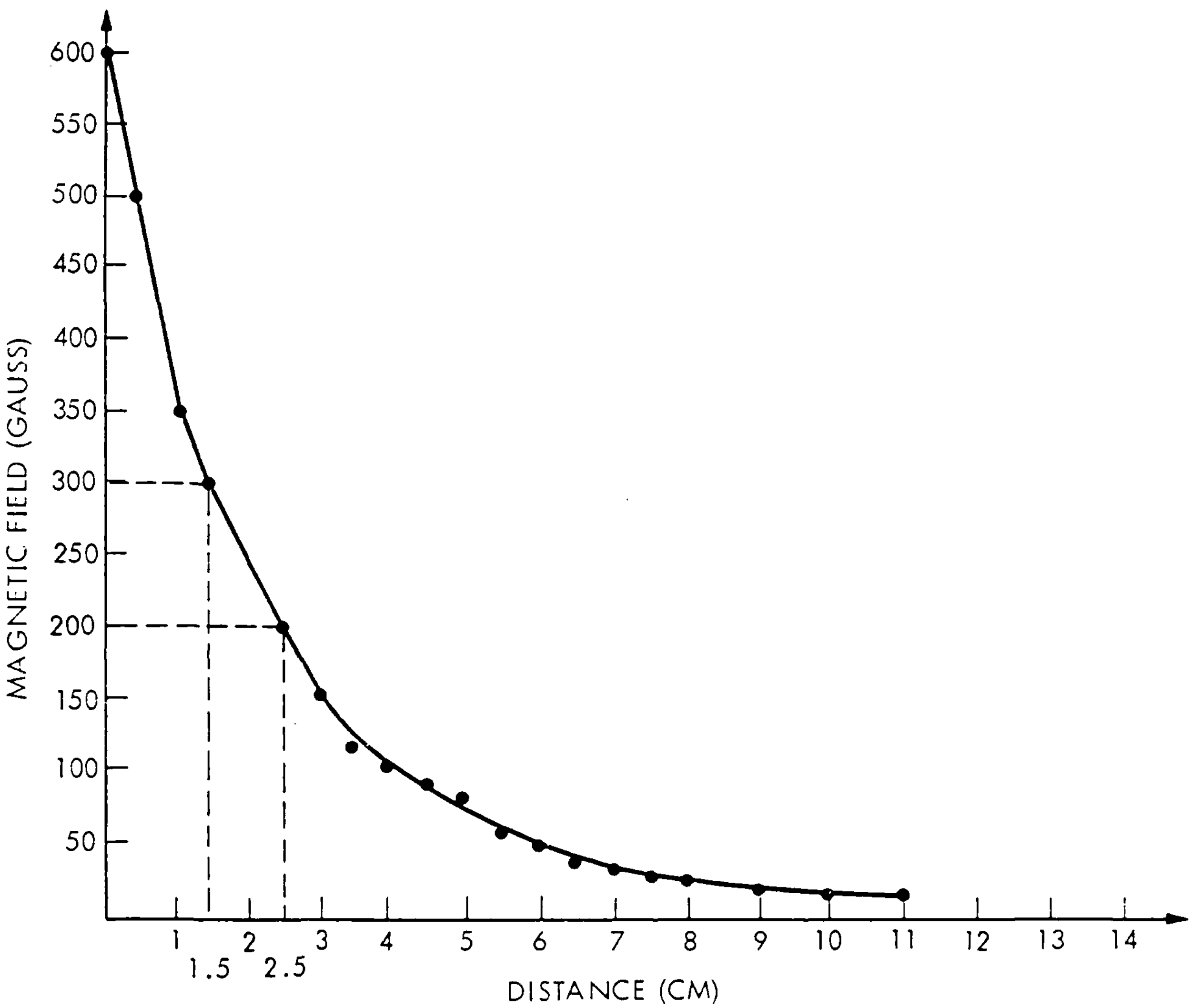


Fig. 7 Axial magnetic field distribution of the ring type permanent magnet

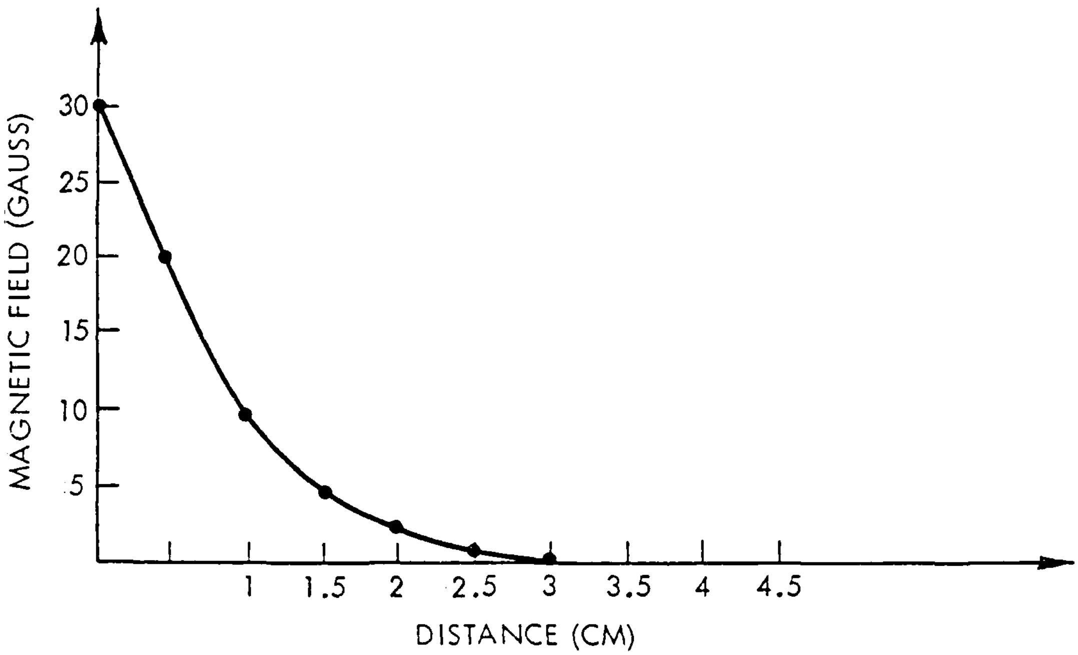


Fig. 8 Axial magnetic field distribution measured along the axis of symmetry of the the external coil when carrying 500 mA current

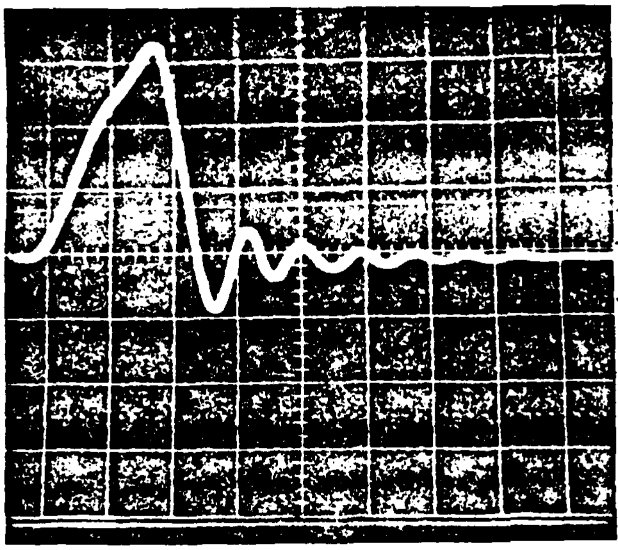


Fig. 9a
X axis: 50 ns/scale
Y axis: 33 ma/scale
bias: 30 gauss

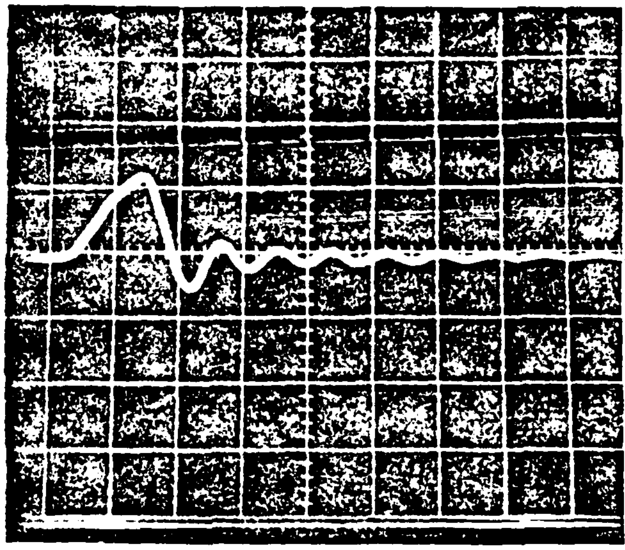


Fig. 9b
X axis: 50 ns/scale
Y axis: 17 ma/scale
bias: 180 gauss

Fig. 9a, 9b Current pulse required to switch a single pixel. Switching threshold: 200 gauss

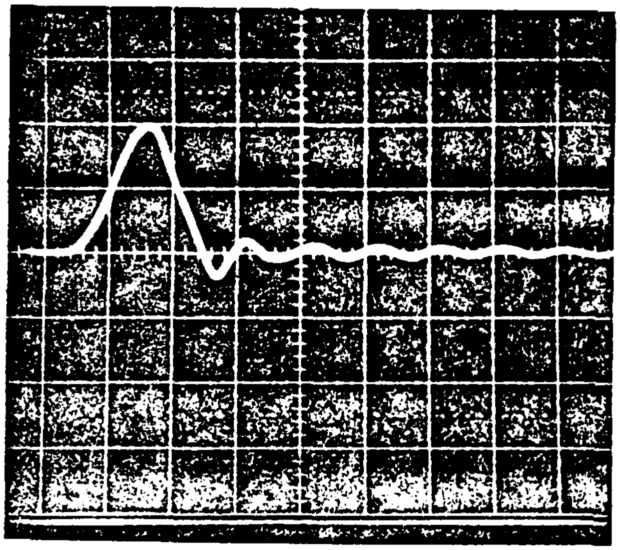


Fig. 9c
X axis: 50 ns/scale
Y axis: 33 ma/scale
bias: 110 gauss

Fig. 9c Current pulse required to switch single pixel at the condition that the driver circuit still can erase all pixels at once by external coil activation only. Switching threshold: 200 gauss

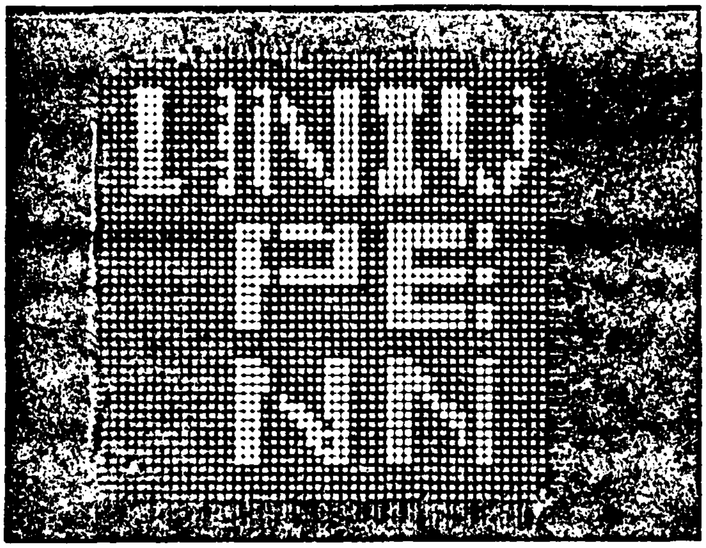


Fig. 10 The pattern written on a 48 X 48 MOSLM. Missing portions of the pattern are due to imperfection of this particular device.

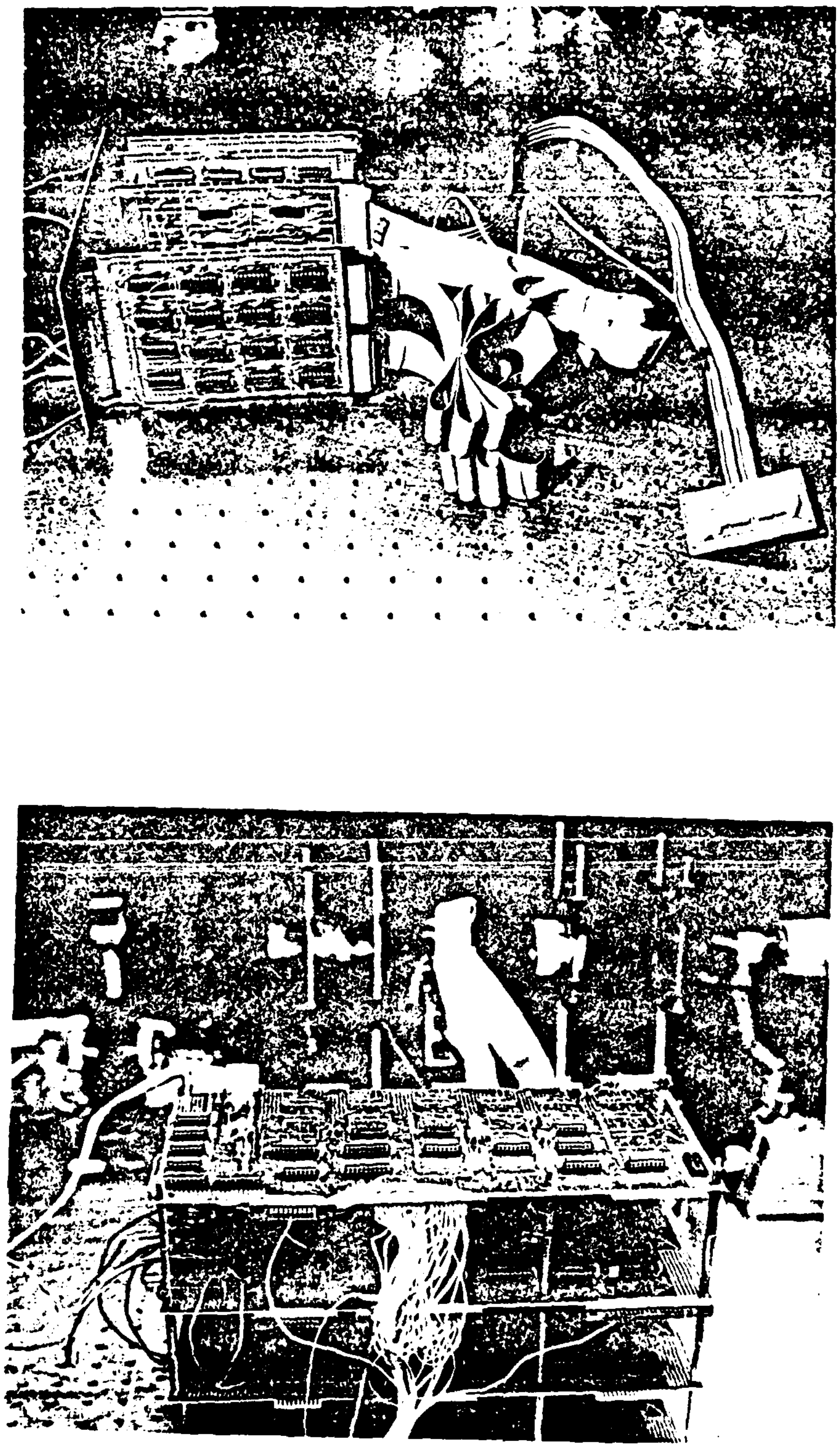


Fig. 11 Pictorial view of driver circuit. (Top) serial addressing scheme (Bottom) staggered addressing scheme

END
DATE
FILMED
5-88
DTIC